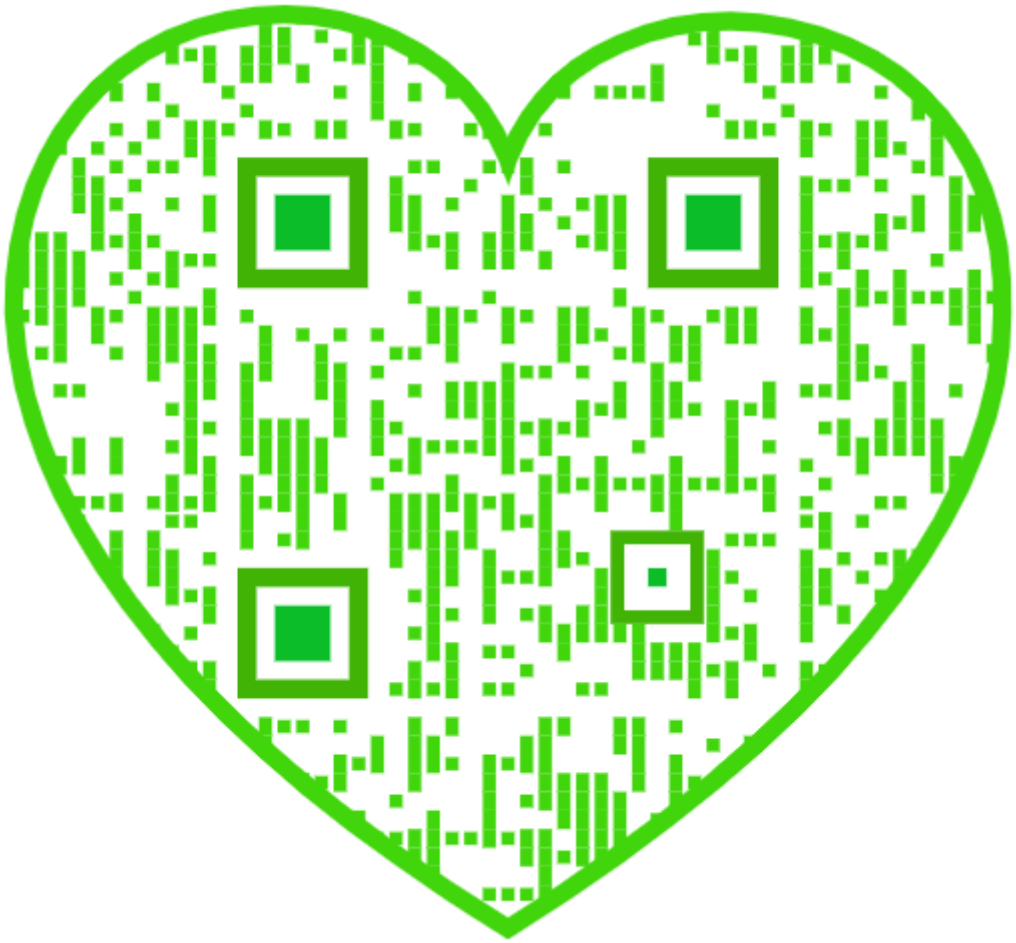
INTRODUCTION TO RADIO ASTRONOMY

R. C. JENNISON

RADIO ASTRONOMY

JENNISON

NEWNES



INTRODUCTION TO

RADIO ASTRONOMY

By R. C. Jennison
B.Sc., Ph.D., F.R.A.S.

This introduction to Radio Astronomy has been written in an attempt to bridge the gap between the textbook and the too-popular treatment of an exciting and rapidly advancing science. It is a book that contains a little more on the techniques of Radio Astronomy than may be found elsewhere and is for those who are not specialists but who none-the-less seek after some real understanding of the subject. A little mathematics has been introduced, especially in the last chapter, but the verbal description of the end products will enable those who shy away from mathematics to glean the significance of the account, and return to the theory if and when they feel inclined.

18/-
NET

Titles available in the NEWNES INTRODUCTORY SERIES are:—

INTRODUCTION TO ALGEBRA
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO ANATOMY
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO BOTANY
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO CALCULUS
15s. cut flush, 21s. cased.

INTRODUCTION TO CHEMISTRY
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO MATHEMATICS
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO MECHANICS
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO PHYSICS
12s. 6d. cut flush, 18s. cased.

INTRODUCTION TO PSYCHOLOGY
10s. 6d. cut flush, 15s. cased.

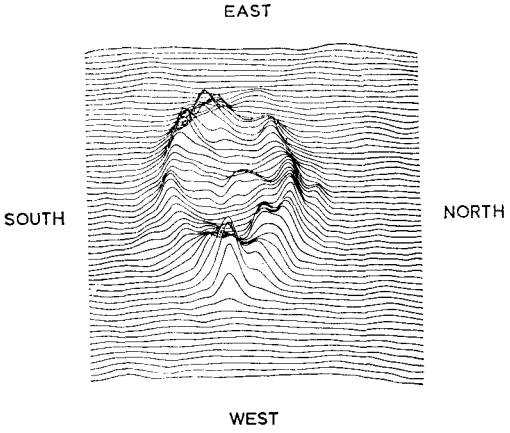
INTRODUCTION TO RADAR AND RADAR TECHNIQUES
10s. 6d. cut flush, 15s. cased.

INTRODUCTION TO RADIO ASTRONOMY
12s. 6d. cut flush, 18s. cased.

FROM ALL BOOKSELLERS

A fully descriptive leaflet is available on request from the publishers,
GEORGE NEWNES LTD., Tower House, Southampton Street, London, W.C.2.

RADIO ASTRONOMY



Frontispiece: Profiles of the radio source Cassiopeia A as presented by an automatic curve plotter. From the first results of the Cambridge supersynthesis survey by Ryle, Elsmore and Neville. This beautiful synthesis of the brightness distribution of the source at 1400 Mc/s was released shortly after this book was received by the publishers, too late for a discussion to be included in the text. Note the peaks of emission around the bright edge (limb) and the gap in the ring a little to the north of East, roughly coincident in position with the earlier observations of a flare (see Chapter Five)

(By permission of the Editors of *Nature*, and Professor M. Ryle)

Introduction to

RADIO ASTRONOMY

by

ROGER C. JENNISON

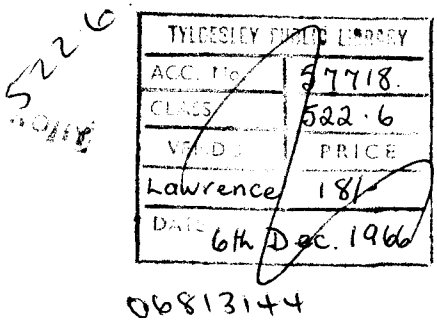
B.Sc., Ph.D., F.R.A.S.

Professor of Physical Electronics,
University of Kent at Canterbury
lately
Senior Lecturer in Radio Astronomy,
Jodrell Bank

NEWNES : LONDON

© R. C. JENNISON 1966

First published 1966



Printed in Great Britain by Butler & Tanner Ltd, Frome and London, for George Newnes Ltd, Tower House, Southampton Street, London, W.C.2

CONTENTS

		Page
	FRONTISPIECE	
	PREFACE	vii
Chapter		
1	Introduction	1
2	The Basic Tools of the Trade	5
3	The Radio Sun	17
4	The Moon and Planets	29
5	Galactic Radio Emissions and the First Radio Stars	45
6	Spectral Measurements	61
7	The Extra-galactic Radio Sources	77
8	Tricks of the Trade	100
	APPENDIX	155
	INDEX	156

PLATES

(*Between pages* 80 *and* 81)

1 Spiral Nebula in Coma Berenices
2 Cluster of Extra-galactic Nebulae in Hercules
3 The 200 ft Jodrell Bank Radio Telescope
4 Interferometer Outstation 100 miles from Jodrell Bank
5 Two 90 ft Paraboloids forming an Interferometer
6 A Solar Outburst
7 Planet Jupiter showing the Giant Red Spot
8 The Crab Nebula in Taurus
9 Nebula associated with the Radio Source Cygnus A
10 The Galaxy NGC 5128 in Centaurus
11 Galaxy associated with the Radio Source Virgo A
12 The 'Quasar' 3C 286

To Jean

PREFACE

Radio Astronomy is a fascinating science. From the most ancient times man has looked with awe upon the wonders of the heavens and has striven to unravel the mysteries and the motions of the objects which he could see and to ponder and postulate in science, philosophy and poetry upon the reason for it all, and for his own role within it. For countless centuries the Universe was studied with the unaided eye, but then came the telescope and a wonderful new range of depth and detail added to his wonderment and his knowledge. Great new optical telescopes pushed the frontiers further and further into space aided by the photographic plate and coloured with the advent of spectroscopy. With all this astronomy became, in the mid-twentieth century, a science of exactitude but of speculation, for the instruments of astronomy were nearing the limits of the capabilities of the traditional patterns of design.

Into this era was born Radio Astronomy, a virile new science which stemmed from radio engineering but finally became established as a powerful complementary ally to the most ancient of the sciences. It has not detracted from astronomy but given it a new impetus and taken it in hand to heights undreamed of when it entered the foray.

There are a number of popular books on Radio Astronomy which whet the appetite for more detailed treatments and for up-to-date accounts of this rapidly advancing subject. There are very few textbooks on Radio Astronomy and those that there are were written many years ago at a level intended for the specialist in the subject. This little book is an attempt to bridge the gap between the too-popular treatment and the textbook. It is an Introduction to Radio Astronomy for those who are not specialists but who hunger for a little more meat within the trimmings of a popular dish. It is definitely not a textbook and it purposely lacks rigour in some of its pages. A little mathematics has been used, especially in the last chapter, but the end products are verbally described and those who cannot digest the mathematics should be able to glean the significance of the account and return to the theory if and when they have the urge.

The book contains a little more on the techniques of Radio Astronomy than may be found elsewhere. This is treated at

two levels, in Chapter Two under the title *The Basic Tools of the Trade*, in which the reader is introduced to the elements of the observing techniques so that he can better understand the following chapters on observations, and in the final chapter, *Tricks of the Trade*, where the sophisticated techniques are explained at an intermediate level for those who crave for more. The remaining six chapters cover most of the observational aspects of radio and radar astronomy tinted with occasional peeps at a little of the colour behind the scenes.

ROGER C. JENNISON

CHAPTER ONE

INTRODUCTION

In the short space of a few decades, Radio Astronomy has advanced Man's knowledge of the cosmos more than any other single factor since the invention of the optical telescope. A whole new spectrum of radiation is now available, far removed from that studied by the optical astronomers—a spectrum extending over many octaves and still so full of rich surprises that many new and fascinating discoveries will be made even while this book is in the press. It is a young and virile science which, together with space research, has rejuvenated the ancient science of astronomy. Yet astronomy was not its parent science; it was born out of radio engineering and nurtured by a radio amateur before astronomy recognized the kindred spirit at the end of the Second World War.

Despite the foundations of the subject in radio engineering, one usually interprets radio astronomy in the context of the Universe as we see it; before embarking on a discussion of the newer science, let us take a peep at the old.

In this peep at the Universe we shall start with the solar system and a few facts which we all know, and then work our way out to the stars and galaxies beyond.

The Moon revolves around the Earth at a mean distance of about 240,000 miles, and, so, as the velocity of light is 186,000 miles per second, moonlight takes about one and a half seconds to reach us and radar echoes sent back to the Earth from the Moon therefore take nearly 3 seconds in transit. The diameter of the Moon is about 2000 miles and its *angular* diameter is half a degree; in other words, it *appears* about the same size as a halfpenny at a distance of 10 feet. The Moon and the Earth together revolve around the Sun, as do the remainder of the planets. All the planets revolve around the Sun in elliptical orbits. The mean distance from the Sun to the Earth, sometimes referred to as the 'astronomical unit', is about 93,000,000 miles and sunlight takes about 8 minutes to reach us. The diameter of the Sun is 860,000 miles, so that its angular diameter is almost the same as that of the Moon, a pure coincidence that occasionally gives rise to that wonderful phenomenon, a total eclipse. The distance from the Sun to the outermost planet,

Pluto, is approximately 3,675,000,000 miles. The Sun emits radio waves continuously and occasionally explosively. The Moon and the planets are relatively weak radio transmitters but may be received at short wavelengths with fairly large radio telescopes. Jupiter also emits curious bursts at much longer wavelengths and these are easier to detect.

The orbits of the planets are crossed by the much more elongated orbits of the comets and meteor streams. Unless we spend most of our lives star-gazing, most of us are lucky if we see one comet in a lifetime, but meteors may be seen quite

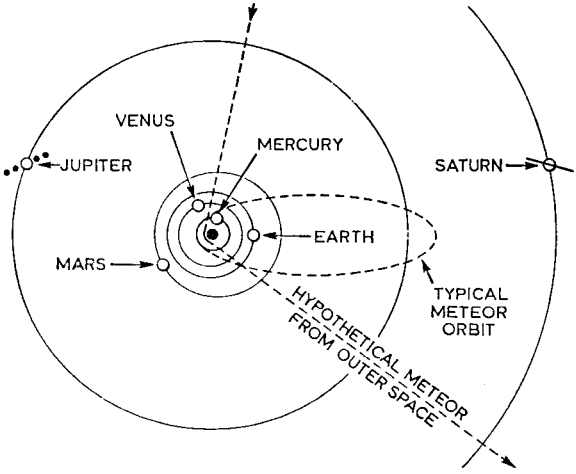


Fig. 1. Orbits in the solar system

frequently, especially on clear nights in the autumn. Most of the meteors seen are tiny pieces of compressed dust, about the size of a pin-head; the dust is distributed in a continuous thin stream all the way round the orbit and it is thought that it may be the débris from disintegrated comets. As the orbits of the meteor streams are narrow and drawn out, each cuts the orbit of the Earth twice during the Earth's yearly voyage round the Sun. On entering the Earth's atmosphere their velocity is of the order of 20 km/s, and they burn up at a height of about 100 km to leave a visible streak by which we see them, and a column of ionized gas which enables us to detect them by radar. It will be apparent from Fig. 1 that the Earth's orbit always intersects that of a particular meteor

shower at the same point, or pair of points. For this reason most meteor showers occur regularly each year, and, furthermore, the trajectories of all the meteors in a particular shower are roughly parallel; as seen by an observer on Earth, therefore, they appear to radiate from a single point in the heavens, called the *radiant* of the meteor shower. All the meteor showers are known by the constellations wherein their radiants appear to be situated; e.g. the Perseids and the Leonids from the constellations of Perseus and Leo respectively. By measuring the velocity of a meteor as it enters the Earth's atmosphere it is possible to compute the shape of its orbit. If the velocity exceeds a certain definite value, known as the parabolic limit, the orbit of the meteor can no longer be an ellipse making a continuous circuit round the Sun, but must be a hyperbola. This means that the particular meteor is not a member of the solar system, and, had it not been intercepted by the Earth, it would have made just one swoop round the Sun and then disappeared again among the stars.

If we journey out beyond the solar system we have to travel about four light years before we encounter the nearest star. Now a light year is the distance that light travels in a year and if we work this out from its speed of 186,000 miles a second it is about 6,000,000,000,000 miles, so that we can visualize the scale if we consider the Sun as the size of a billiard ball in London, the nearest star would then be an object of about the same size in Edinburgh! The separation between the Sun and its neighbours is typical of that between most of the stars in our Galaxy, and yet our Galaxy contains about 100,000,000,000 stars. All these stars are distributed in a flat disc having a spiral structure and a hub at the centre, rather like a gigantic Catherine wheel except that, instead of one spiral arm, the Galaxy probably has two. The solar system is situated in one of these great tentacles about two thirds of the way out from the galactic nucleus, or hub, to the edge of the disc. The actual distance of the Sun from the nucleus is about 25,000 light years. When we look into the sky on a clear night and see the Milky Way we are actually seeing the stars in the plane of the disc; looking to either side of the Milky Way the stars appear less numerous as we are looking outwards through the nearer stars at right angles to the plane of the disc. Why, one may ask, do we not see the bright nucleus of the Galaxy? The answer is that we do, but only at radio wavelengths; at optical wavelengths it is obscured by clouds of dust. These dust clouds are

similar in form, though not in content, to ordinary earth-bound clouds but are on the average about thirty light years across. The distance between the particles of dust may be many miles but the aggregate is sufficient to black out all the stars shining from behind. The dust clouds are confined to a very thin sheet in each spiral arm coincident with the plane of the Galaxy; when we look up at the Milky Way they appear to us as holes in the belt of stars. Similar clouds may be seen in other galaxies when they are viewed edge on, as in Plate 1. Also roughly coincident with the dust clouds are clouds of gas forming a thin sheet running through the Galaxy: this gas is almost entirely hydrogen and has a mean density of about one atom per cubic centimetre, that is about one fifty-million-million-millionth of the density of air, yet it is one of the chief sources of emission of radio waves.

Our own Galaxy with its myriads of stars is just one of millions of similar galaxies scattered through space; they stretch to the limit of the observable universe, and there may be many more than the most intensive searches with the largest telescope have yet revealed. Though in some parts of the sky these galaxies occur in clusters, such as in Plate 2, they move about to a large extent independently of each other, and as the distance between typical galaxies is not many times the size of each, collisions between them may sometimes occur. The radio source in Cygnus, the second brightest of the radio 'stars', was once thought to be an example of such a catastrophe, but the theory ran into many difficulties and it is now thought that although such collisions are almost commonplace in the cosmos, they have little to do with the radio stars.

The distance to the Cygnus radio source is over 500,000,000 light years. The distance to the Andromeda nebula, the nearest of the external galaxies, is about 2,000,000 light years and its diameter is about 130,000 light years. The most distant galaxy observed at the time of writing this chapter is a nebula in the direction of the constellation of Boötes and is 4,500,000,000 light years away. The light from this galaxy, as with other distant galaxies, is very much redder than that from our neighbours. This reddening increases with distance and is accompanied by a proportionate shift in the wavelength of the spectral lines which cannot be accounted for by simple scattering or filtering of the light in space. It may result from a recessional Doppler shift such as would occur in an expanding universe.

CHAPTER TWO

THE BASIC TOOLS OF THE TRADE

The signals which a radio astronomer detects have the character of random electrical 'noise', but he rarely listens to the sound of the signals and it is seldom that they are of such a type and intensity that he can arrange his apparatus to give a virtually instantaneous two-dimensional picture of the source. Usually he is content to study the temporal changes of the total radiation within the bandwidth and beam width of his apparatus as it remains fixed upon, or slowly scans through the source in one dimension. The operation of scanning is greatly simplified by the fact that the Earth rotates on its axis, so that a fixed aerial upon the surface of the Earth will scan a complete circle of the sky in the course of a day. This is known as scanning in *right ascension*. Scanning in *declination* may be achieved by tilting the aerial system in a plane at right angles to the scan in right ascension, that is, by tilting it towards or away from the pole of the Earth.

Some of the radio telescopes used in present-day radio astronomy are rather sophisticated systems operating on novel but elegant principles which will be discussed in Chapter Eight. Other telescopes are extremely simple. The first detailed sky survey was performed by Grote Reber in 1940 with a home-made paraboloid in his back garden. The paraboloid used as a radio telescope is just the radio analogue of Newton's reflecting telescope of three hundred years ago, from which both the 200 in. optical telescope at Mount Palomar and the 250 ft radio telescope at Jodrell Bank (Plate 3) are direct descendants.

For many experiments in radio astronomy the telescope is even more rudimentary than a Newtonian reflector and becomes just a simple aerial system. Radio signals from solar outbursts (Plate 6) are the strongest cosmic signals that are received on Earth and these may be detected with exceedingly simple apparatus.

Radio noise signals from the *quiet* Sun decrease in intensity at longer wavelengths while the irregular enhanced radiation from solar flares and other manifestations of the *disturbed* Sun increases in intensity as the wavelength is increased. In order to detect most of the solar radio emission anywhere in the

range of frequencies from 30 to 200 Mc/s a simple dipole or 'H' aerial connected to a stable receiver will suffice. The band width of the receiver should be as large as possible, consistent with the avoidance of man-made interference; the bandwidth and overall performance of a good television set is adequate for most purposes. The detector circuit should feed a smoothing circuit having a time constant of about one second in order to 'integrate' the noise output prior to passing the signal to a pen recorder or other device suitable for forming a permanent record (Fig. 2).

If the detector has a square law characteristic then the increments of d.c. output are directly proportional to the increments of noise power at the aerial, but the noise ripple on the d.c. output, tending to mask very small changes of level, is inversely dependent upon the square root of the product of the

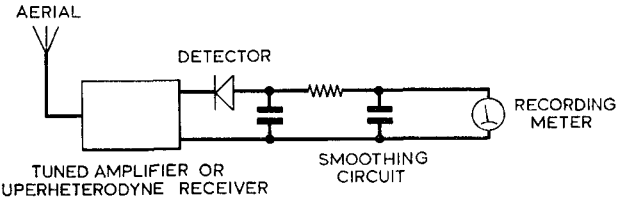


Fig. 2. A simple radio astronomy receiver

pre-detector bandwidth and the integration time (the time constant of the smoothing circuit on the recorder). To detect very small changes of noise entering the aerial one can therefore use either a very large aerial to increase the increments at the input, or a broad band receiver having a large time constant on the detector, provided that the receiver is sufficiently stable and free from interference.

A similar aerial and receiver working at low frequencies will give a clearly visible increase in output as the Milky Way moves through the aerial beam, but as the beam width of a dipole is very wide, the receiver itself must stay constant in gain for a very long time, practically a whole day. The detection of the radiation from the quiet Sun places similar requirements on the stability of the receiver if only a dipole aerial is used, though the very much stronger bursts of noise from sunspots and solar flares, when they occur, may be detected easily on an unstabilized receiver. It is not difficult to stabilize the power to a transistor receiver but if vacuum tubes are used, both the high

and low tension supplies should be stabilized for all this work. There are several circuit arrangements for the receiving equipment which only give an output for noise actually entering the aerial, and this alleviates the problem of stability. In one of these methods, shown in Fig. 3, the input of the receiver is continually switched between the aerial and a noise diode,

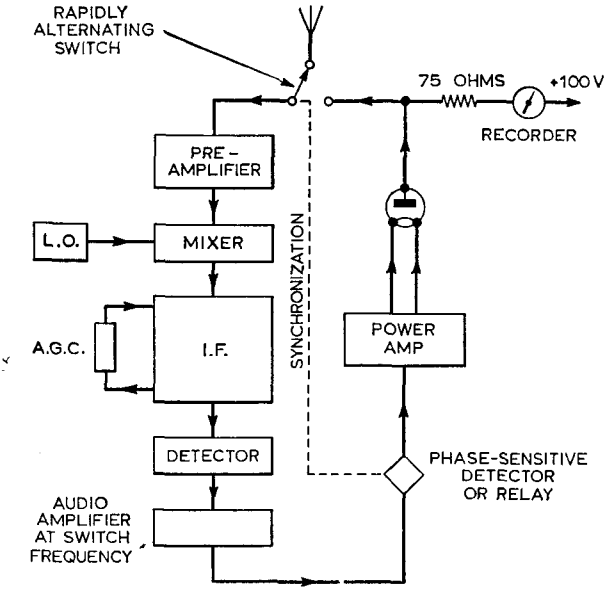


Fig. 3. A servo controlled receiver in which the signal is rapidly compared with that from a noise diode
(Ryle and Vonberg, following a principle first used by R. H. Dicke)

while the output is switched in synchronism to feed a corresponding difference signal back to the filament of the noise diode; the output of the noise diode is therefore always automatically maintained at the same level as the noise entering from the aerial, and the level of the latter may be directly observed on a milliammeter in the anode circuit of the diode. An entirely different method of obtaining an output from only the noise signals entering the aerial is to use a phase-switched, or a phase-rotated, interferometer. We shall return to this method later.

Very simple aerial systems and equipment are used for the detection of the ionospheric scintillations which occasionally modulate the signals reaching the earth from the radio stars.

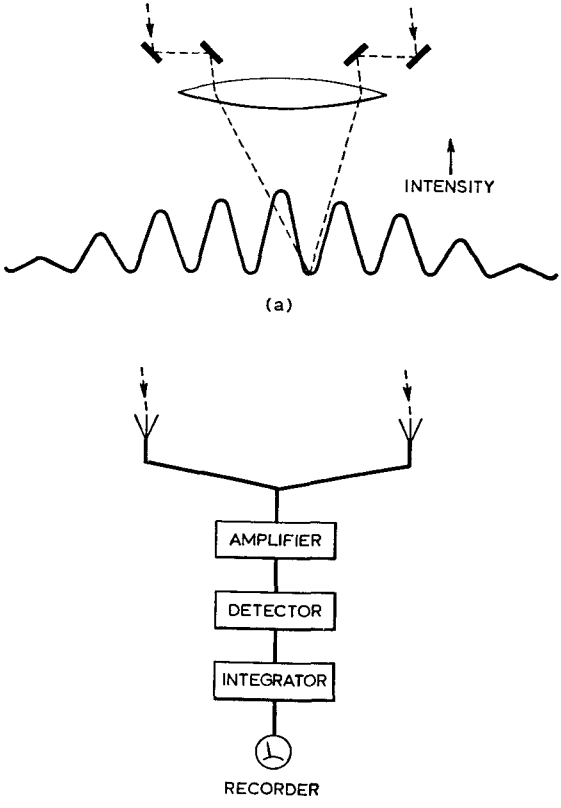


Fig. 4. The Michelson stellar interferometer
(a) optical technique
(b) radio analogue

These scintillations correspond to the twinkling of ordinary stars and are due to the inhomogeneity of the upper layers of the Earth's atmosphere. If, say, a single Yagi aerial, that is the familiar type resembling the backbone of a fish that is mounted

on many a chimney-top for television reception, is connected to a simple receiver tuned to any frequency between 30 and 100 Mc/s, the radiation from the Galaxy will be detected and the noise output of the receiver will be found to rise during the transit of the plane of the Galaxy through the aerial beam. About once or twice a week it may be observed that there also occurs a rapidly fluctuating signal lasting for a time comparable with the motion of a celestial body through the aerial beam. These signals will be due to the radio stars in Cygnus and Cassiopeia, the two most intense of the radio sources, but the fluctuations are caused by the F layer of the ionosphere, which modulates the signals from the radio sources. By comparing the records of twinkling received at two or three stations spaced about a mile apart it is possible to determine the size and shape of the irregularities in the ionosphere and also the speed at which they move. This simple experimental technique has contributed greatly to our knowledge of the ionosphere in recent years.

The large radio telescopes which have recently been constructed have the great virtue of enormous collecting power, but they are not essential for the study of many of the problems in radio astronomy. The fine beam width, or high resolving power, of the very large aerial systems may be obtained and even exceeded with only two dipoles if they are connected as an interferometer, that is to say they are connected by equal lengths of feeder cable but are spaced many wavelengths apart (Fig. 4; Plates 4 and 5). This system of two aerials gives a polar diagram consisting of a number of fine lobes and the width of each lobe is half that which would be obtained in the main beam of a single aerial system of such a size that it filled all the space between the two dipoles! The transit of a radio star through an interferometer aerial beam gives, on the pen record of the receiver output, a bump caused by the broad polar diagram of the individual aerials, modulated by very clearly marked fringes as the source passes through the interference pattern of fine lobes (Fig. 5). If the radio star has an angular diameter smaller than the width of these lobes the modulation of the bump by the fringes will be 100 per cent, but if the source is larger than the lobes the modulation will be reduced and may not even appear. The interferometer system therefore is ideally suited, not only for the detection of very small regions of emission, but also for their accurate positioning and the measurement of their shape and size.

A great improvement of the simple interferometer, introduced by Martin Ryle, was to incorporate a switching mechanism whereby a half wavelength of feeder was rapidly and

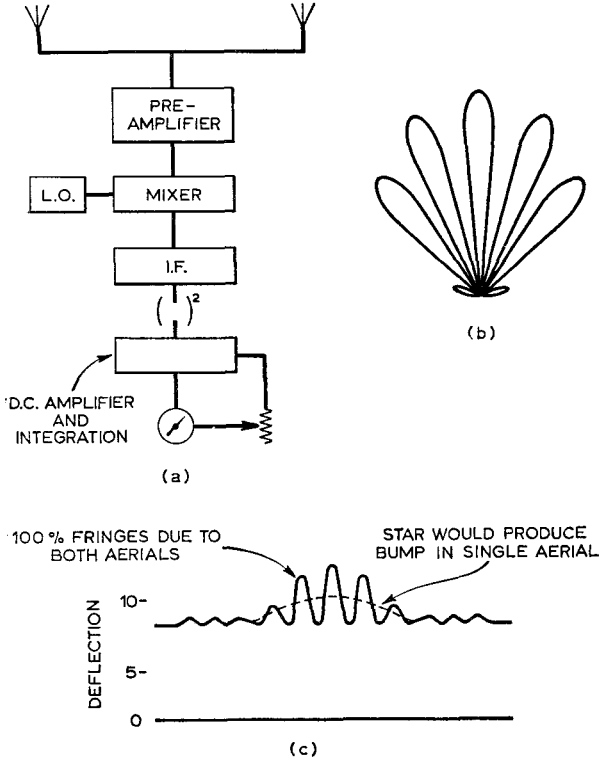


Fig. 5.

(a) Superheterodyne equivalent of Fig. 4

(b) Polar diagram

(c) Output on recorder chart. The bump caused by the star crossing the polar diagram of the individual aerials is modulated by the interference fringes produced by the varying phase difference between the two aerials

periodically inserted in one of the two main feeders, the output of the receiver was then arranged to record only the difference of noise level between these two states by inserting a syn-

chronized relay in series with the meter (Fig. 6). The great advantage of this method is that it is not affected by large variations in the background illumination of the sky; the

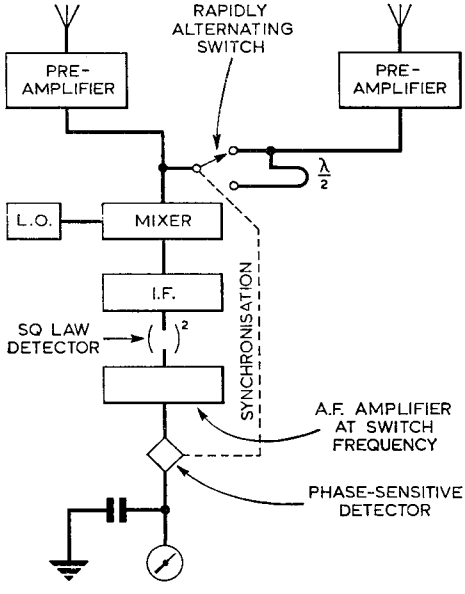


Fig. 6.
(a) The phase-switched interferometer
(b) The output in the presence of a source
The output in the absence of a source is a straight line along the centre of the chart

stability requirements are considerably reduced and the sensitivity of the receiver may be kept constant by using an automatic gain control without appreciably affecting the size of

the fringes in the output. A variation of this technique is to rotate the phase of the signals in one of the two limbs of the interferometer and thus to generate artificial fringes in the presence of radiation from a small source; with this arrangement the natural fringes may be slowed down or speeded up at will, while they may also be produced when the two aerials are on a north–south baseline, a condition which gives no fringes with a simple interferometer. It is permissible to insert separate pre-amplifiers in each aerial lead and to use separate mixer stages so that the two channels may be united at an intermediate frequency, but if such an arrangement is used it is essential to use the same local oscillator to feed both mixers.

A radio link may be used to unite the two channels of an interferometer if the baseline is very long and in such an arrangement it is necessary to transmit back not only the signal in the distant channel but also all oscillator frequencies used to form the finally transmitted signal. These frequencies must then either be applied to the 'home' channel or subtracted from the distant channel at the receiving end in order to preserve the coherence of the signal (see Chapter Eight).

The power in the spectrum of the radiation from the Galaxy increases markedly at the lower frequencies, and depending upon the period in the sunspot cycle and the corresponding state of ionization of the upper atmosphere, the radio noise that it emits can be detected easily on frequencies as low as 10 Mc/s provided that we observe the Galaxy when it is almost overhead. The power spectrum of the individual radio 'stars' also rises with decreasing frequency, though not as sharply as that from the Galaxy, and Cassiopeia A and Cygnus A may also be detected with relatively simple apparatus at frequencies below 200 Mc/s. In the British Isles these two radio 'stars', the strongest in the sky, transit at upper culmination, i.e. are nearest overhead, about 5 degrees north and 15 degrees south of the zenith respectively. The time of day when they are in this position varies with the time of year, for the stars complete one circuit round the pole star in one siderial day. The siderial day is 4 minutes shorter than the ordinary day so that the time of transit of a particular star moves 4 minutes earlier every day. At the right time of the year, one could therefore listen to an ordinary television receiver when the programme transmitters were 'off the air' and, pointing the aerial at the sky, one could receive the signals from these sources. By connecting a pen recorder in place of the input to the cathode-ray

tube the signals could easily be detected, but might not be recognized unless the aerial were connected as an interferometer.

An entirely different type of signal from that already described may also be detected from the Galaxy. This is the hydrogen line which occupies only a very narrow frequency band

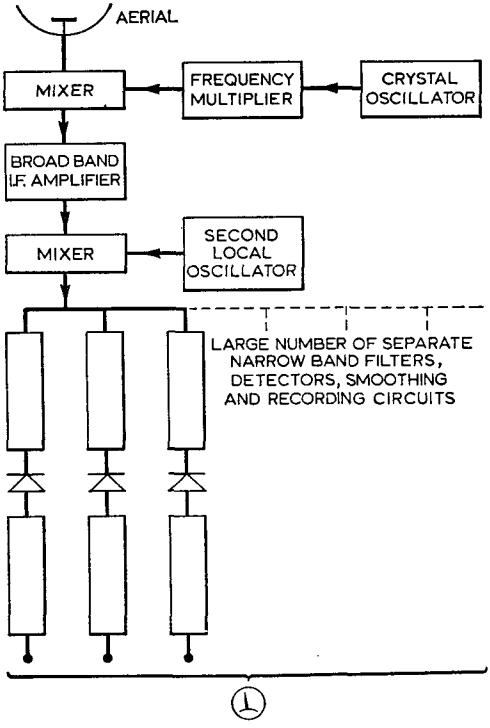
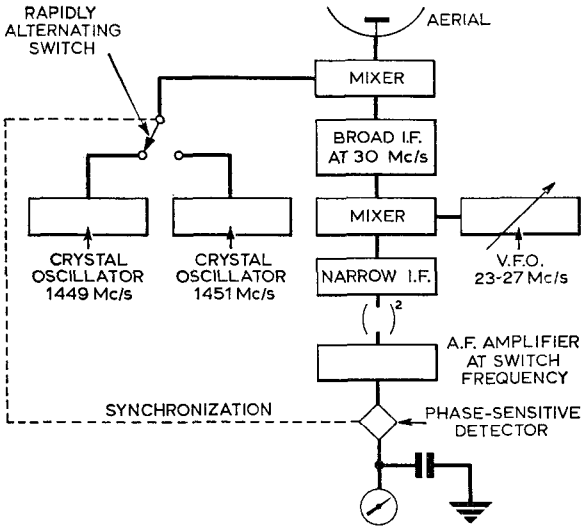


Fig. 7. Simple multi-channel hydrogen line spectrometer receiver

centred on 1420 Mc/s. This frequency is emitted by the neutral hydrogen atoms in the Galaxy when the magnetic fields of the proton and electron in each atom change from being parallel to anti-parallel. The very important point about this signal is that it gives an estimate of the speed at which different parts of the Galaxy are moving, and from our knowledge of the



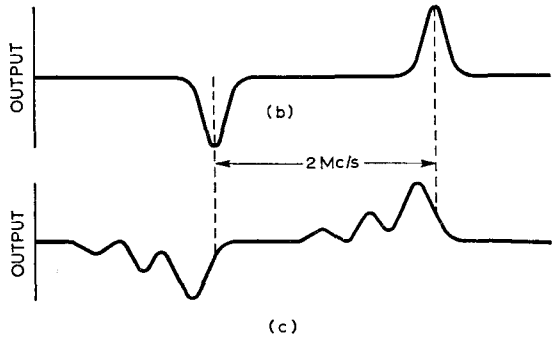


Fig. 8.

(a) Switched comparison receiver

(b) Output as variable frequency oscillator (v.f.o.) is tuned through the line at 1420 Mc/s

(c) Typical output when line is complex

approximate shape and dynamics of the Galaxy we are able to fill in the accurate details of its structure. The reason why we are able to measure the velocities of the transmitting regions from this signal is that the motion of the transmitter shifts the frequency at which it is received, a phenomenon known as the Doppler shift. This is exactly the same phenomenon as that which causes the note of a locomotive whistle apparently to change in pitch as the train passes through the station; if we measure the change of pitch we can calculate the speed of the train.

The most usual form of the apparatus used to study the hydrogen line is called a spectrometer. Two types of hydrogen line spectrometer are shown in Figs. 7 and 8. Figure 7 shows a simple multi-channel system, while Fig. 8 consists of a receiver in which the noise in two frequency bands separated by a few megacycles is compared in level. Both channels are tracked together over a frequency range from about 1418 Mc/s to 1422 Mc/s. The line appears first in one channel and then in the other and the output of the receiver is first deflected in a positive and then a negative direction, in each case tracing the contour of the line (Fig. 8 (b) and (c)). The minimum aerial requirement for satisfactory results from the hydrogen line is probably a 10 ft paraboloid. This paraboloid would have to be accurate to a tolerance of one inch but could be made very simply by stretching chicken mesh across 3 or 4 hoops rigidly fastened to a frame in the form of two Vs at right angles, joined together at the apex.

Spectrometers are also used for studying the radiation from the active Sun. These spectrometers scan over a much greater range of frequency than those used for investigating the hydrogen line. The technique most frequently used is to gang-tune the radio frequency and local oscillator circuits of a superheterodyne receiver and to drive this tuning system from a motor so that it covers a frequency range of almost two to one, say from 30 to 50 Mc/s. When the scan reaches 50 Mc/s, switches are brought into operation which immediately cause the receiver to continue sweeping over the greater part of the next octave, say from 50 to 90 Mc/s, and so on up the frequency scale as far as may be required. Figure 9 shows a technique used in Australia with which it is possible to scan continuously from 200 to 2000 Mc/s. It is usual with these spectrometers to record the output of the receiver as a function of time and frequency by moving a 35 mm film past a

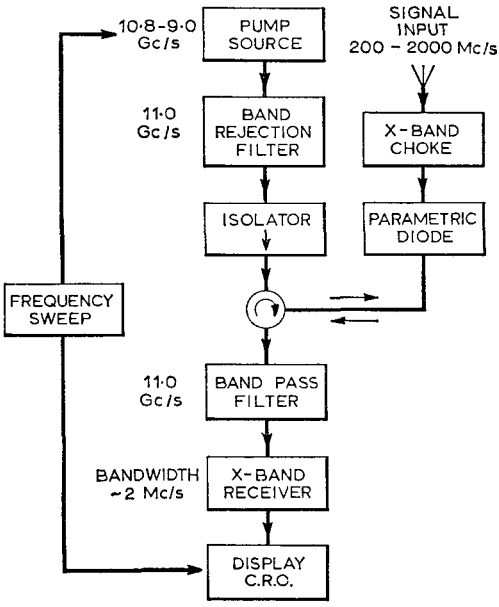


Fig. 9. An Australian solar spectrometer with large dynamic range (200 to 2000 Mc/s). A moving film is passed in front of the cathode-ray oscilloscope (c.r.o.) in order to obtain a permanent record of the variations at different frequencies as a function of time, as in Fig. 12

small cathode-ray tube, the beam of which is intensity modulated by the strength of the signal and is scanned across the width of the moving film in synchronism with the frequency sweep.

CHAPTER THREE

THE RADIO SUN

The strongest of the radio signals reaching the Earth from outer space emanate from the Sun when it is active or disturbed with sunspots.

In view of the intensity of these signals it may seem curious that they were not the first cosmic signals to be recognized, but their discovery came about 10 years after Jansky detected the radiation from the Galaxy and was a result of the operational research conducted by J. S. Hey into the jamming and interference which affected wartime radar receivers.

Solar activity follows an 11 year cycle but the outbursts, when they occur, are transient phenomena and are fairly randomly distributed in time within any particular year, there being statistically many more at times of sunspot maximum than sunspot minimum. Had there been intense sunspot activity in the early nineteen thirties it is likely that Jansky would have recognized the origin of the signals in the course of his investigation of cosmic static.

THE QUIET SUN

It is now known that there are other radiations from the Sun in addition to those which come from active areas. Even when the Sun is 'quiet' or undisturbed by sunspots it radiates radio waves just because it is a large hot object. All hot objects emit electromagnetic radiation to an extent which depends upon their temperature. Most of the radiation occurs in the visible and infra red spectrum but there is also radiation tailing right through the infra red into the longer wavelengths of radio waves. The total quantity of radiation from an object increases as its temperature increases and at the same time the peak of the spectrum of its radiation moves further into the shorter wavelengths. If one compares a red hot poker with one which has been raised to 'white' heat in a furnace one will obviously feel much more heat radiated from the white hot poker; while if one examines both through a blue filter, only the white poker will be seen to be giving blue light. Both pokers will radiate very weak radio signals (random noise) but the hotter poker

will give the stronger signals. This radiation is in accordance with the well known laws of black body radiation.

The Sun does not behave in quite the simple manner of a red hot poker. At visible wavelengths it appears to be an object about half a degree in diameter and at a temperature of approximately 6000°K. At centimetre wavelengths it appears a few per cent bigger and has a temperature of about 10,000°K. At metre wavelengths both the size and the temperature increase very considerably so that it appears to be about a degree in diameter and to have a temperature of a million degrees. The way in which the temperature and diameter increase together enables us to explain the curious behaviour as a result of the radiation from different levels in the solar corona which must get *hotter* as it becomes less dense further away from the Sun. This corona is highly ionized and the laws of refraction and electromagnetic propagation prevent us from seeing to (or radiation reaching us from) those layers of the corona where the electron density exceeds a certain critical density which varies with the wavelength or frequency of our observations. The critical frequency at any level in the solar atmosphere is also known as the *plasma frequency* of the ionized gas,

$$f_p = \mathrm{e}\sqrt{\frac{N}{\pi m}}$$

where f_p is the plasma frequency,
N is the electron density,
e is the charge on the electron, $4{\cdot}8 \times 10^{-10}$ e.s.u., and
m is the mass of the electron, $9{\cdot}1 \times 10^{-28}$ g.

Radiation at the frequency f_p can only escape from the level in the corona where the electron density has thinned out to be equal to or less than the appropriate value of N given by the formula. A very similar phenomenon occurs on the Earth where the ionized upper layers of the Earth's atmosphere prevent the escape of radio signals from terrestrial transmitters whose frequency is below that of the plasma frequency of the ionospheric layers. It is of no avail transmitting to the Moon from the Earth at a frequency of 1 Mc (300 m) for the signal would be refracted by the ionosphere and bent back to Earth, whereas a signal at 100 Mc is well above the plasma frequency and can escape without impediment.

With the quiet Sun the radiation does not, of course, come from radio transmitters such as those on the Earth's surface, but from the so-called free-free transitions as the electrons in

the ionized gas pass close to the protons. The speed of the free electrons increases with the temperature of the gas; as an electron approaches a proton it is attracted by the opposite charge and its course is altered so that it swings round in a hyperbolic orbit. Any such change in the speed or direction of motion of an electron means that the electron is accelerated and it is an established law of physics that an accelerated charge radiates electromagnetic energy. All the electrons in the solar corona are continually being accelerated in a random manner by their encounters with the protons, and the result is the emission, absorption and re-emission of radiation over a continuous range of frequencies limited only by the laws of thermodynamics and propagation which we have already considered. The result is that the short, centimetre wavelengths originate almost entirely in the relatively dense chromosphere while the metre and decametre waves appear from the higher levels in the solar corona.

The theory of the radiation from the quiet Sun predicted that at shorter wavelengths the edge of the disc should appear to radiate more strongly than the centre because the extra thickness of gas encountered by tangential rays increases its opacity and therefore its ability to absorb and emit radiation. This increase of brightness towards the edge is known as limb brightening. No predictions of limb brightening were made for the longer wavelength radiation from the corona because no positive temperature gradient was expected and because the electron collisions in the coronal gas are such that it reflects and refracts the waves more efficiently than it absorbs them. These predictions were vindicated in the measurements made with interferometer equipment at a variety of wavelengths. Limb brightening was observed at centimetre and decimetre wavelengths while at metre wavelengths the corona showed limb darkening together with a radial asymmetry which was not predicted in the theory but which might have been expected from the marked asymmetry of the optical emission of the corona which is observable at times of total eclipse.

The strength of the signals from the quiet Sun at different wavelengths is shown in Fig. 10. This graph gives in watts the flux that falls on each square metre of an aerial directed at the Sun for each cycle per second of the bandwidth of the receiver. Thus at a wavelength of 3 m the flux is 10^{-22} W/m^2 for each unit of bandwidth. If the bandwidth of the receiver is 1 Mc and the area of the aerial is 100 m^2 the power received would be

$10^{-22} \times 10^6$ W for each square metre of aerial, giving a total signal of $10^{-22} \times 10^6 \times 10^2$ W assuming that all the radiation falling upon the aerial reached the receiver. In practice the aerial collects only half the radiation because the incoming radiation contains waves with two components of polarization and the aerial detects only one. There may also be a further loss resulting from the inefficiency of the aerial and its associated cables.

In order to assess how large this signal may be it should be compared with the noise generated in the receiving equipment.

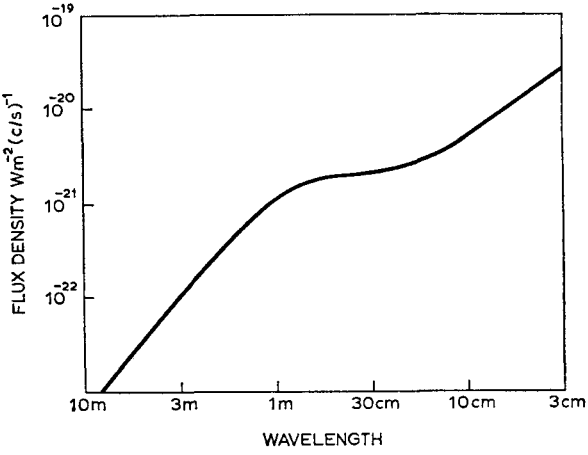


Fig. 10. The spectrum of the quiet Sun

If this is an unsophisticated receiver without a parametric or maser pre-amplifier it will probably have a noise factor of two or three, which means that it will behave as though it were at a temperature of about 600° to 1000°K, and the noise generated in the receiver is simply

$$kTB \text{ Watts},$$

where k is Boltzmann's constant ($1{\cdot}38 \times 10^{-23}$ joules per degree),

T is the temperature (see footnote on p. 29), and

B is the bandwidth in c/s.

Thus in the case we have considered, if we take a receiver with a bandwidth of 1 Mc, and an input 'temperature' of 1000°K,

the receiver noise power is $1{\cdot}38 \times 10^{-20} \times 10^6$ W which we may compare with the useful signal of $\frac{1}{2} \times 10^{-22} \times 10^6 \times 10^2$ W. The factor of 10^6, representing the bandwidth, cancels out in this calculation and we are left with a signal to noise ratio of about one third. This signal would be very clearly detectable if the receiver were reasonably stable and a small amount of smoothing were applied to the output of the receiver. (A simple integration network of a series resistance and parallel capacity giving a time constant of a few seconds would be typical.)

The emission of signals from the corona of the Sun may be detected from the Earth on low frequencies out to heights of about two solar radii. Beyond this height radio measurements have been made of the electron density in the outer corona by a very different technique. Radio signals from certain very distant radio sources are occulted by the corona at those times in the year when the Earth, Sun and radio source are in appropriate positions in space. The refraction of the ionized gas in the outer corona causes a bending and scattering of the rays from the source and this may be detected in carefully controlled experiments made throughout a period embracing the entire occultation. Hewish, Vitkevitch and Slee have all contributed to this field and the effect of the corona may now be detected out to heights of nearly a hundred solar radii.

THE SLOWLY-VARYING COMPONENT

When there are sunspots visible on the Sun's disc an enhancement of the radiation may be detected above the level associated with the quiet Sun. This enhancement may last for a period of days or even months but if months it is likely to disappear and reappear in synchronism with the 27 day rotation period of the Sun. The fact that the reception of these enhanced signals depends upon solar rotation suggests that they come from localized areas on the Sun's disc, and very accurate directional observations, particularly by Covington and Christiansen, have shown that the radiation emanates from the dense condensations associated with 'plage' areas near sunspot groups. The centre of the radiating region, about the size of the underlying sunspot, emits strongly circularly polarized radiation while radiation with less polarization is emitted from the wider area corresponding to the visible calcium plage. The height of this emission is not easy to determine but approximate measurements, made by studying the displacement in position between the centre of the radio emission and the optical

sunspot for enhancements observed at various positions on the Sun's disc, indicate that the heights extend up to 50,000 km.

The mechanism for the generation of the enhanced slowly varying component of solar radiation is not yet understood. Free-free transitions such as those responsible for the radiation from the quiet Sun would not explain the change in slope of the spectrum of the flux of the radiations at wavelengths between 5 and 10 cm. A clue to an alternative process lies in the fact that the radiation shows strong circular polarization. This implies the presence of a strong magnetic field, typically about 500 gauss and it has been suggested that some of the radiation is caused by electrons gyrating in this magnetic field, radiating low harmonics of the gyro frequency,

$$\frac{eH}{2\pi mc},$$

where H is the magnetic field,
c is the velocity of light, and
e/m is the ratio of charge to mass of the electron.

The plage regions on the Sun are also the regions which appear brightest when X-ray photographs of the Sun are taken by equipment flown in rockets, and the radiation from these regions has a pronounced effect upon the ionization of the upper layers of the Earth's atmosphere.

THE ACTIVE SUN

In February 1942 the wartime radar stations in Britain were affected by strong interference which reduced their sensitivity to aircraft by enhancing the noise level of the receivers so that the echoes were lost within the 'grass' or random noise on the radar displays. The Germans had tried a number of ways of jamming the radar system and it was possible that this was a new form of jamming, but if so it was extremely subtle. The reports were investigated by Dr. J. S. Hey of the British Army Operational Research Unit, and he showed that the enemy on this occasion was the Sun. The findings were originally classified but were publicly reported immediately after the war. Appleton and Hey also published a paper in which they showed that there was a correlation between the strongest enhancements of noise and optical observations of flares on the Sun.

Solar flares are most frequently observed at times of maximum sunspot activity, when sometimes as many as several

small flares an hour may be recorded. The same is true of the transient radio outbursts from the Sun; the likelihood of an outburst occurring depends upon the time of observation in the 11 year sunspot cycle.

The solar flare appears optically as a sudden brightening of the light from a small area on the Sun, usually close to a sunspot group. If one looks at the brightening with a spectroscope one finds that it occurs particularly to the spectral line from excited hydrogen known as the Hα line. The enhancement of brightness occurs suddenly and then decays more slowly over a period of many minutes and occasionally nearly an hour. Very large flares are rare but often show structure in the form of filaments. Jet-like surges may be seen either brightly, in emission, or darkly, in absorption against the bright background. Sometimes huge arched prominences (such as that shown in Plate 6) erupt into the solar atmosphere. The optical emission is usually accompanied by ultra violet and X-ray emission which reaches the Earth 8 minutes later and causes excess ionization in the Earth's ionosphere. This effect, known as a Sudden Ionospheric Disturbance, or S.I.D., may cause a temporary disruption of short wave radio communications on Earth.

High energy cosmic ray particles are emitted on the occasion of large flares. The highest energy particles arrive at the Earth about 10 minutes after the flare, while the lower energy particles may take several hours. About 30 hours after the flare it is not unusual for a corpuscular stream to envelop the Earth. This is a jet or stream of charged particles which travels out from the Sun at a speed of typically 1500 km/s and upon arrival in the vicinity of the Earth interacts with the Earth's magnetic field, and the particles trapped therein (the van Allen Belts), causing a magnetic storm, a decrease in the cosmic ray flux and aurorae.

The radio emissions from the active Sun may be recorded over the whole range of wavelength from a few metres to a few centimetres but they change in character over the spectral range and some of the individual disturbances are confined to one end of the spectrum (Fig. 11). The signals picked up at the longer wavelengths are on the whole stronger and more ragged than those at the shorter wavelengths. When it is possible to record the signals simultaneously on two widely spaced frequencies, it is found that a disturbance usually occurs first at the higher frequency and appears a minute or so later on the

lower frequency. The changing character and time delays of the signals in different parts of the spectrum make it very desirable to study a large range of wavelengths at virtually the same

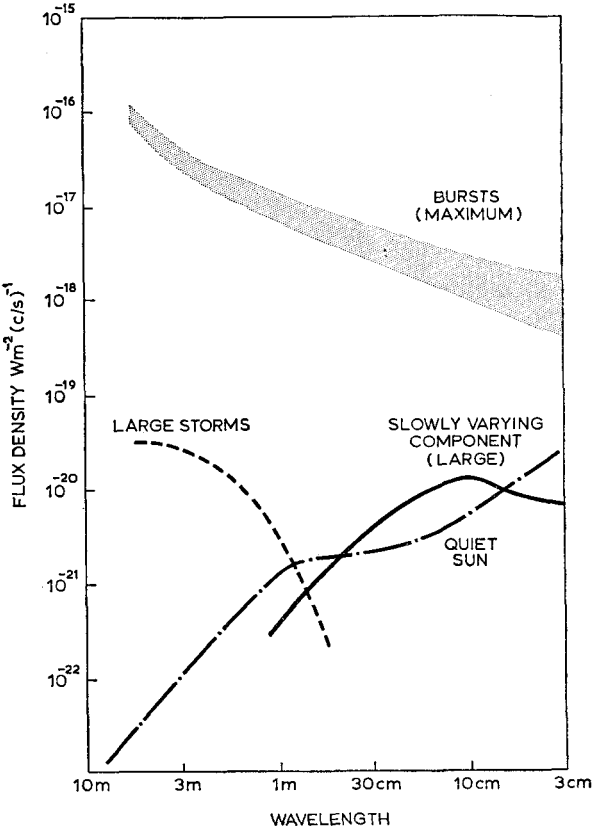


Fig. 11. The spectra of the active Sun
(After J. P. Wild)

time. This is made possible by the use of a spectrometer which could be a large number of separate receiving channels all operating simultaneously but which is much more usually a single receiver arranged to tune rapidly and continuously over

a very large range by a motor on the tuning shaft. The output of the receiver is applied as intensity modulation to a cathode-ray tube, the beam of which is scanned in synchronism with the motor drive. A further clock motor winds a 35 mm film past the cathode-ray tube so that the finally exposed film carries a succession of images of the swept signal in which the strength of the signal appears as the brightness of the image as in Fig. 12. This type of spectrometer was developed by Paul Wild in Australia and is now accepted as a basic part of the equipment of any solar observatory.

With the aid of spectrometers the radio astronomers have been able to classify the radio emissions into a number of

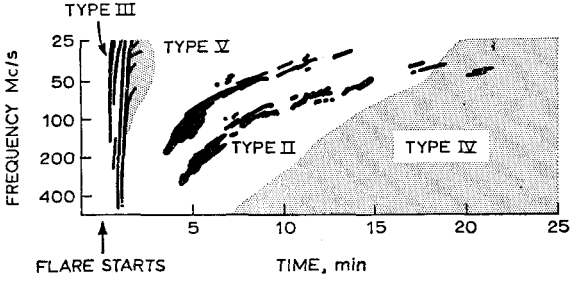


Fig. 12. Idealized dynamic solar spectrum following an outburst (After J. P. Wild)

different types according to their spectral and temporal characteristics supplemented by polarization measurements. Outbursts of spectral type I are narrow band bursts lasting less than a minute, usually circularly polarized and often superimposed upon a slowly varying background of broad band enhanced radiation. The bursts are seen only on metre wavelengths where they follow each other in rapid succession to form a continuous storm which may last for several days while the part of the Sun responsible for the radiation, the region above an active sunspot, moves across the centre of the Sun's visible disc. This implies that the type I and noise storm radiation is directively beamed and this can be accounted for by the effects of refraction in the solar atmosphere. It is more difficult to account for the actual mechanisms of generation of the noise storm radiation. The circular polarization implies the presence of a magnetic field and this is to be expected close to a sunspot, but it does not necessarily imply that the radiation must be of

the cyclotron or synchrotron type (see Chapter Five), and most current theories favour an origin in plasma oscillations.

Outbursts of spectral type II are violent events lasting from 5 to 30 minutes. The radiation begins at high frequencies and drifts slowly to the lower frequencies. At any one instant the radiation occupies a band of frequencies a tenth or more of an octave wide, often with structure within this band. Usually a fair copy of the structure appears simultaneously an octave higher, indicating the presence of pronounced harmonics. The emission from type II bursts is attributed to plasma oscillations associated with a disturbance which moves outward through the solar atmosphere at a velocity of the order of 1000 km/s. This could be consistent with emission occurring as a magnetohydrodynamic shock wave, caused by the explosion of the flare, passes outwards through the corona.

Bursts of spectral type III are more common than those of type II. They occur in groups lasting about a minute, usually accompanying the smallest of the optical flares, and detected at metre wavelengths occasionally showing pairs of bands spaced an octave apart. These bands sweep through the frequency spectrum from high to low frequencies much more rapidly than the type II bursts. It is thought that they may be caused by a stream of electrons ejected from the flare rushing outwards through the solar corona and exciting plasma oscillations of progressively lower frequency as the density of the corona decreases on the outward journey. The rate of the frequency drift combined with the known variation of density of the corona makes it possible to compute the speed of the electron stream and this turns out to be typically between a quarter and half the speed of light. A more detailed investigation shows that the motion appears to occur along the coronal streamers where the density is somewhat higher than the average density of the corona.

It is not unusual for type III bursts to be accompanied by a smoother burst at centimetre wavelengths. These centimetre bursts, usually called *microwave early bursts* closely resemble the bursts in the X-ray spectrum detected by balloon and rocket-borne apparatus. The origin of the X-rays is thought to be in the free-free transitions of the electrons (Bremsstrahlung), as the fast electrons pass through the chromosphere, and it is likely that the associated microwave burst is produced by synchrotron radiation as the electrons pass through the strong magnetic field in the active region.

TABLE 1. SOLAR RADIO EMISSIONS

T ye	*Duration*	*Spectrum*	*Polarization*	*Location and related phenomena*
Quiet Sun	Continuous	All wavelengths	Unpolarized	Thermal radiation from chromosphere or corona
Condensations	Weeks (Slowly varying)	All wavelengths	Centre only polarized	Hot dense plages in corona
Noise storms	Hours or days	Long wavelengths	Circular	Sunspot radiation
Type I burst	Tens of seconds	Long wavelengths	Circular	Near sunspots (storm burst)
Type II burst	Minutes	Dynamic spectrum, moving slowly from short to long waves	Unpolarized	Sometimes associated with ejection of corpuscular streams
Type III burst	Seconds	Dynamic spectrum moving rapidly from short to long waves	Unpolarized	Associated with flares and probably with ejection of cosmic ray particles
Type IV burst	Minutes or hours	Noise storm over many octaves	Ultimately circular	Occasionally follows large flare
U burst	Seconds	Dynamic spectrum, wavelength rapidly increases then reverses	—	Similar to type III
Type V burst	Minutes	Long wavelengths	—	Occasionally follows groups of type III

A large radio burst of type II is usually preceded, several minutes beforehand, by the complete sequence of a microwave early burst and a group of several short and sharp type III bursts. If the type II burst is a very large event it is usually followed by a very long period of enhanced radiation known as type IV. This radiation extends simultaneously over nearly the whole spectrum and it has been found that the source may move away from the flare region at a very high speed (typically 1000 km/s). A type IV event has the remarkable property, unlike the type II and type III features, of emitting radiation over many octaves from the position that it occupies at any one time. This has lead to the belief that the mechanism of radiation is the synchrotron process caused by electrons spiralling at almost the speed of light while trapped into a magnetic field (see Chapter Five). In the final phase of this continuum emission it becomes markedly circularly polarized and settles down over the original active centre. In this final stage the synchrotron mechanism may give way to cyclotron and plasma radiation.

So far we have discussed four types of events associated with solar eruptions. Some of these, in particular type IV, are further divided and given suffixes to indicate particular components or features. A variation of the type III burst is the 'U' burst first observed by Alan Maxwell. This has a spectral characteristic which reverses as though the disturbance returns to denser regions. At the time of writing there is one further major disturbance, a type V which is a burst of continuum radiation lasting about a minute which occasionally follows groups of type III bursts.

A summary of solar radio emissions is shown in Table 1.

CHAPTER FOUR

THE MOON AND PLANETS

Any hot dense body emits radiation in the radio spectrum with an intensity appropriate to its temperature according to the well known laws of thermal radiation. Provided that the body is not too far away and is sufficiently large, it is possible for this relatively weak thermal radiation to be detected even though it may represent a temperature much lower than the equivalent noise temperature of the receiving equipment.

The signal strength or flux density of the thermal radiation received from a hot body increases with the solid angle subtended by the body at the aerial and decreases according to the square of the wavelength of reception. Thus upon the assumption that the black body temperatures* of the Moon and planets are likely to lie in the range from about 50 to 1000°K, it would be reasonable to expect that the object which subtends by far the greatest solid angle, namely the Moon, would be the easiest object to detect. From the wavelength dependence of the flux density it also follows that the measurement could be performed using a relatively small aerial at very short wavelengths.

Thus it was that in 1946 Dicke and Beringer were able to detect the radio emission from the Moon using a delightfully conceived stable receiver operating from a 3 ft diameter paraboloid at a wavelength of 1·25 cm. Their original work has been followed by many detailed investigations on a variety of wavelengths and a comprehensive picture of the radio Moon now emerges.

THE MOON

At wavelengths of a few centimetres the radiation from the Moon varies during a lunar month in a roughly sinusoidal fashion, but these variations do not show the extremes of hot and cold that are indicated by infra red measurements. The maximum of the radio emission does not occur at full Moon but about three and a half days later, whereas the infra red

* It is usual to express the temperature in degrees Kelvin (°K) which uses similar units to the centigrade scale, but starts at absolute zero, i.e., −273°C. On this scale, therefore, water boils at 373°K and freezes at 273°K.

radiation reaches a maximum simultaneously with the visible full Moon. The radio emission from the Moon has been studied during a number of lunar eclipses but no change of the signal level has been observed. This is consistent with the fact that the duration of the eclipse is very much shorter than a lunation and the effect of the temperature change in the eclipse is therefore smoothed out.

At metre and decimetre wavelengths the sinusoidal variation of the radio emission over monthly periods disappears and only a constant temperature is recorded. Measurements by different observers and at a variety of wavelengths give somewhat different values for the mean temperatures averaged over the lunar disc, varying between about 180°K and 270°K.

The behaviour of the Moon in the radio frequency spectrum can be explained if the visible and infra red emissions come from the surface, while the radio emission originates at some depth beneath the surface where the temperature variation, due to solar radiation, is reduced in amplitude and shifted in phase by the thermal properties of the surface material. The longer the wavelength of reception, the deeper will be the seat of the emission beneath the surface of the Moon. It therefore follows that the thermal and electrical properties of the surface layers of the Moon can be studied by the use of radio techniques. The results of the radio observations are consistent with a lunar surface material having similar thermal and electrical properties to those of dust or porous rock such as pumice.

PLANETARY RADIO EMISSIONS

The first radio emission to be detected from a planet was a surprise in many ways. Typical of a number of scientific discoveries such as the well known story of the discovery of penicillin, it was the outcome of the brilliant following up of an anomaly in the expected behaviour of an experiment.

Burke and Franklin of the Carnegie Institution in Washington were in 1955 observing the sky with a long aerial having quite a small beam width at a longer wavelength than most other experimenters were using at the time. The aerial was tuned to 20 Mc/s and the beam width was only about 2°. It was not an easy aerial to adjust and it was kept for some time trained on an interesting part of the sky through which the Sun, the radio source in Taurus (the Crab Nebula) and the nebula IC 443 all transited. The fact that the planets might also pass through the beam, if considered at all at this stage, was dismissed

as inconsequential, for it was unlikely that their thermal radiation would be received, and at 20 Mc/s it would be exceedingly weak. As the days and weeks passed, Burke and Franklin were disturbed by the presence of ragged and noisy interference on their records which appeared at first to occur at roughly the same time every day and, though it looked and sounded rather like the interference from the ignition system of a motor vehicle, it did not coincide with the operation of any vehicles or machinery in the neighbourhood. After a while it became clear that it was not, after all, occurring at quite the same time every day, nor was it keeping time with either the milking of the Maryland cows or the transits of the distant stars. Nevertheless it had to have some real origin. They looked along the direction of the aerial beam when the interference was at a peak and there, behold, was Jupiter. Cross checking their records it was clear that Jupiter was always in the beam when the signals were received, but not one radio astronomer in the world would have dared to suppose that it could send out such curious signals.

The discovery is even more remarkable when a final tally is taken of the coincidences. It is now known that Jupiter only emits these signals in quite a narrow band centred on about 20 Mc/s—just the frequency that Burke and Franklin were using. Their aerial had a narrow beam width which enabled them to recognize the variation in time of the phenomenon and the direction from which it was coming, and they happened to have this aerial pointing in the direction through which Jupiter would pass. But for this curious combination of circumstances and the admirable insight of Burke and Franklin, many a crackle picked up on 20 Mc/s today would be dismissed as wretched interference, and one of the fascinating facets of our own solar system might still cry out, unheard and unobserved.

The announcement of Burke and Franklin's discovery was followed by feverish efforts to investigate the signals in more detail and to re-evaluate earlier records which might show the mysterious effect. It was found that they had been received in Australia on a frequency of 18 Mc/s for some years previously but nobody had attached any importance to the dirty little clusters of spikes that had appeared on the records. Further research showed that the signals were remarkably confined to a narrow part of the frequency spectrum and that they were circularly polarized.

These impulsive radio signals from Jupiter show a periodicity

of 9 h 55 min 30 sec. As this is closely similar to the rotation period of Jupiter about its own axis, it was natural to look for a correlation between the times of reception of the radiation and any markings upon the planet which might at the same time be presented in the direction of the Earth. If Jupiter is studied with an optical telescope it is observed that its atmosphere appears to be divided into a number of bands, parallel to the equator of the planet (Plate 7). These bands exhibit temporary and sometimes quasi-permanent markings which make it possible to measure their rates of rotation, and it is found that the period at the equator is 9 h 50 min 30 sec, while at the poles the rotation takes 9 h 55 min 40 sec. Thus the periodicity of the radio bursts agrees with that of a belt near one of the poles of the planet and from the precise time of reception of each group of bursts it is possible to establish both the Jovian latitude and longitude of the source with reasonable accuracy.

Despite the reasonably accurate fixing of the radiation region by these methods, it is found that it does not appear to coincide with any particular feature in the planetary atmosphere. It may, however, be significant that there is at about the same latitude, though not at the same longitude, the most conspicuous of all the markings on the planet Jupiter—the giant Red Spot. This remarkable feature is a blemish which was first observed in 1878 when it was brick red in colour and 30,000 miles long by 7000 miles wide, big enough to contain the Earth. It remained very prominent for many years and then gradually diminished, becoming more circular and losing its red colour until it is now seen as a notch or hollow in the edge of the great southern belt. From a study of early drawings of the planet it has been possible to identify this same hollow as far back as 1831 and the red spot itself since 1859. It is difficult to reconcile the behaviour of the spot with a feature on the solid surface of the planet, for it has markedly changed its period since it was first observed and it occasionally speeds up and slips back again in position as it is overtaken by a dark region in a belt which rotates at a slightly faster rate just south of the spot.

At very short radio wavelengths Jupiter behaves in a rational manner and the radiation detected at 3 cm corresponds to a temperature of about 150°K which is consistent with the infra red measurements. At decimetre wavelengths the behaviour of the planet is again anomalous for the apparent temperature of the planet rises to a value which is quite inconsistent with a thermal origin for the radiation (Fig. 13). At still longer wave-

lengths the radiation appears to decrease again until eventually one reaches the onset of the curious burst like radiation centred on about 15 metres.

The enhanced radiation in the decimetre band exhibits a

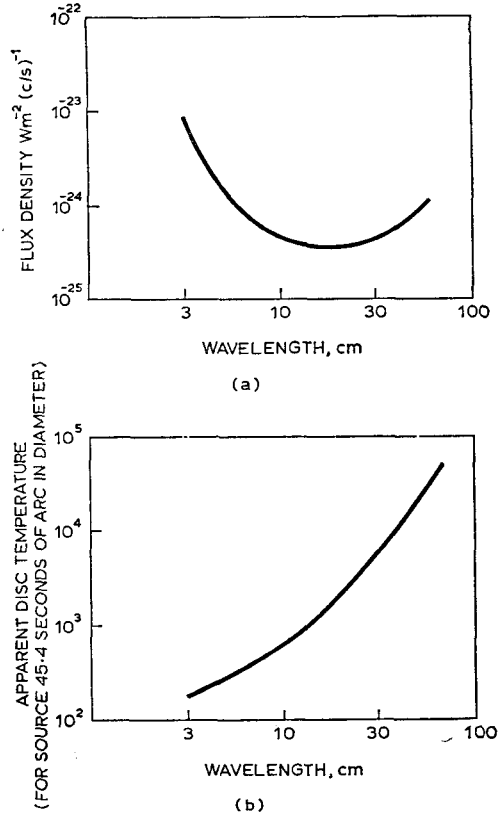


Fig. 13. The decimeter spectrum of Jupiter

long period variability although the extent and regularity or otherwise of this variability has not yet been established. There may also be a shorter period variability consistent with the rotation period of the planet's higher latitudes but observers differ in their findings on this score. The radiation at 31 cm

is found to be about 30 per cent linearly polarized with the electric vector of the radiation approximately parallel to the Jovian equator. At this wavelength the source of the radiation appears to be about 3 times larger than Jupiter in the direction of the equator.

There have been many attempts to explain the enhanced decimetre radiation of Jupiter. The most likely source may be synchrotron radiation originating in radiation belts similar to the van Allen belts of the Earth. If this is so the magnetic field must be about 5 gauss and the density of the electrons must be about 100,000 times the density of the particles in the van Allen belts. The observed radio spectrum is not in accord with the energy distribution of the particles in the Earth's radiation belt and it is at present debatable whether the required energy distribution could exist in the Jovian belts. An alternative suggestion is that the radiation comes from non-relativistic electrons spiralling in the magnetic field of Jupiter. If this is so a field greater than 1200 gauss would be required at Jupiter's magnetic poles.

Observation of the two planets nearer to the Sun than the Earth, Venus and Mercury, is somewhat complicated by the fact that when they are closest to the Earth, and therefore in principle easiest to detect, they can, in fact, only be observed with large aerials at very short wavelengths. This is because they are closest to the Earth at inferior conjunction, when they are aligned between the Earth and the Sun. The Sun is quite a powerful radio emitter so that it would virtually blind the radio telescope if it were received in the same main beam, and even if it were only received in the side lobes of the aerial system it would still constitute a powerful signal which would embarrass the measurements. A very fine aerial beam width is therefore required for this measurement. In view of the fact that, if the spectrum of radiation from these two planets is similar to thermal radiation, the signals will be strongest at high frequencies, it is apparent that they can best be detected by a large aerial operating at short wavelengths.

At the time of writing, Mercury has only been observed at 3·45 cm. The measurement was performed by the University of Michigan using an 85 ft radio telescope connected to a maser. The mean aerial temperature observed was 0·05°K which corresponds to a mean disc temperature of the planet of 400°K. This temperature must be interpreted with care for it corresponds to the temperature which the whole surface of the

planet, if presented in the direction of the Earth, would need to have to give the observed aerial temperature. Optically only a crescent would be seen so that it is highly probable that the figure of 400°K is the mean of the large area of cool planet combined with the hot crescent receiving the Sun's rays. Mercury always presents the same aspect to the Sun,* for it rotates on its own axis once a year, and the side nearest to the Sun is likely to be roasting at about 1000°K, while the far side may freeze at between 25° and 50°K.

The behaviour of Venus as an emitter of radio waves is rather interesting. In the millimetre spectrum the observed temperature of the planet is about 350 to 400°K which is slightly higher than the temperature observed at infra red wavelengths. In the centimetre spectrum the temperature rises considerably and is about 600°K at 10 cm. This behaviour is inconsistent with thermal radiation from a black body but it might be explained if the centimetre radiation originated on the surface of the planet, while the millimetre and infra red radiation came from the Venusian clouds. This would imply that the surface of Venus is much hotter than we might have expected while the dense clouds which mask it, and prevent us optically even seeing the surface of Venus, are relatively cool. Nevertheless a temperature of 600°K for the surface requires some explanation.

It is unlikely that the centimetre emission is of non-thermal origin as it is difficult to account for the observed spectrum and absence of polarization in this way. Two alternative theories are more promising. One of these is that Venus has a very dense ionosphere and that this radiates thermally. It would require an electron density of 10^9/cc uniformly filling an atmospheric layer 1000 km thick to account for the observations. This may be compared with an electron density of the order of 10^6/cc in layers about 100 km thick in the Earth's ionosphere.

An older and possibly more plausible theory is that the atmosphere of Venus contains considerable quantities of carbon dioxide and water vapour which trap the infra red radiation and produce a high temperature at the surface. In order to produce this hothouse effect nearly all the solar radiation must be trapped and very little can escape by radiation from the clouds. Carbon dioxide alone cannot account for the observed

* Very recent measurements have yielded the surprising result that Mercury probably rotates on its own axis once every 59 days. These measurements were made with radar by Pettengill and Dyce of the Lincoln Laboratories in the U.S.A.

temperatures. One suggestion was that the effect of the carbon dioxide is augmented by dust in the atmosphere but a more likely explanation is that water vapour is responsible. If Venus rotates on its axis only about once an orbit then a partial pressure of water vapour of about 1 g/cm^2 would enhance the the effect of the carbon dioxide sufficiently to account for the observations. In the humid and predominantly carbon dioxide atmosphere at the surface of the planet it is interesting to conjecture if life of any sort could exist. If indeed the surface temperature is really 600°K it is very difficult to imagine the existence of life as we know it in any of its earthly forms, for the temperature corresponds to the melting point of lead.

Of the remaining planets only Saturn and Mars have been conclusively detected. A search has been made for radiation from Uranus using the 210 ft Australian radio telescope at a wavelength of 11 cm. These observations set an upper limit to the temperature of Uranus as 320°K and served only to show that the planet does not emit strong non-thermal radiation at this wavelength.

Mars has been observed only at wavelengths of about 3 cm. This work was done by the Naval Research Laboratories near Washington and yielded a disc temperature of 211 $\pm$ 20°K. The temperature measured in the infra red region is about 250°K so that the radio temperature is about 40° less. It is likely that this phenomenon is similar to that which occurs on the Moon and that the radio emission originates at a small depth beneath the surface of the planet. The radio temperature on the planet is therefore likely to correspond to the mean temperature of a hypothetical Martian wine cellar, little affected by day and night but very cold by earthly standards and not recommended for earthly wines. The infra red temperature on the other hand corresponds to the temperature of the sunlit surface. Subject to correction for albedo, Mars appears to offer a rather chilly welcome.

Saturn has been observed at 3·75 cm by Drake and Ewen, and more recently at 3·4 cm by the University of Michigan, using a ruby maser on their 84 ft telescope. The observed aerial temperature was 0·095°K corresponding to a disc temperature of 106 $\pm$ 21°K after making a 2 per cent correction for atmospheric absorption. This value is in reasonable agreement with the temperature of 123°K measured in the infra red by Menzel, Coblentz and Lampland in 1926. It is not yet known if Saturn has an additional non-thermal spectrum at longer wavelengths

in the manner of Jupiter. It is also uncertain whether the rings of Saturn emit appreciable radiation.

LUNAR AND PLANETARY RADAR

The solar system contains the only astronomical objects which are likely to be within the range of radar systems in the foreseeable future. Radar reflections from the Moon were first obtained by Bay in Hungary in 1946 with considerable difficulty at the time. At the time of writing, Moon radar is a relatively simple task and sophisticated surveys of the Moon can be performed by this means. Planetary radar is currently limited* to Venus but has yielded valuable information on the rotation of the planet and on the size of the astronomical unit.

If a radar signal is transmitted from the Earth to a distant object, the power received by that object will depend upon the power (P) of the transmitter, the gain of the transmitting aerial (G) and the distance of the object from the Earth (R). If the object is a disc of radius a, the total power that it receives will be

$$\left(\frac{PG}{4\pi R^2}\right)(\pi a^2) \text{ Watts}$$

Some of this radiation falling on the planet will be absorbed by the planet where it is ultimately transformed to a minute quantity of heat. The remaining power is reflected back into space and a small part of it will return in the direction of the earth. *En route* to the Earth the intensity falling on a given area again diminishes according to the inverse square law so that on final reception at the Earth the strength of the signal received by an aerial of effective area A_e is

$$\left(\frac{PG}{4\pi R^2}\right)(\pi a^2 \eta)\left(\frac{A_e}{R^2}\right)$$

where η is a factor depending upon the efficiency with which the target returns the radiation. If the same aerial is used for transmission and reception G and A are simply related,

$$G = \frac{4\pi A_e}{\lambda^2}$$

(see Chapter Eight). The strength of the received signal is therefore

$$\frac{PA_e^2}{\lambda^2} \cdot \frac{\pi a^2 \eta}{R^4}.$$

The parameters on the left hand side of this expression are in

* See footnote on p. 35.

the hands of the experimenter for they represent the transmitter power, the size of the aerial and the wavelength upon which it is used. The parameters on the right hand side all depend upon the astronomical object, its actual size, its efficiency as a reflector and its distance from the Earth. It is the distance from the Earth which most severely limits the use of radar, for it appears in the expression inversely to the fourth power; an increase of ten times in the distance of an object would reduce its echo to one ten-thousandth.

In order to determine whether or not the received power constitutes a detectable signal, it is necessary to compare it with the noise power present in the receiving system. The narrower the bandwidth of the receiver, the less will be the noise, but if the radar is pulsed the bandwidth must not be reduced too far. The received pulse power will be diminished if the receiver bandwidth is less than the natural bandwidth of the pulse, which, by Fourier's Theorem, corresponds to the reciprocal of the duration of the pulse. In continuous wave radar this limit does not apply but the bandwidth has to be sufficient to receive the side bands of the returned signal produced by the rotation of the planet which causes a Doppler shift modulation. The optimum bandwidth in both cases is obtained by matching the receiver bandwidth to the frequency spread of the returned signal, that is the width of the pulse in the first case and the Doppler spread in the second.

The study of the Moon by radar made great strides in the late 1950s and many lessons were learned in the early days which have now resulted in reliable and sophisticated techniques. It was found that the received signals showed a pronounced fading and this fading appeared to fall into two categories. One of these was quite rapid while the other occurred over much longer periods. The longer period fading was very troublesome and initially very puzzling for it did not appear to be associated with any known properties of the Moon. It was studied extensively by W. A. S. Murray and others and was eventually shown to be caused by the ionosphere of the Earth. The effect is due to a phenomenon known as Faraday rotation. In the early days the aerials used for Moon echoes were linearly polarized and they could therefore receive a signal at full strength if it were polarized in the same direction, whereas they would not receive it at all if the polarization were at right angles. A homely example is that from a television transmitter. Television signals are usually (not always) transmitted with

vertical polarization and if there are no objects about to scatter the radiation and rotate the plane of polarization, the signals will be received with a vertical aerial but not at all with a horizontal aerial. If the radio waves pass through a region where there are lots of free electrons and a magnetic field the plane of polarization is rotated. If the region is extensive it may spiral round many times. This phenomenon is known as the Faraday effect. When the waves were emitted from the old lunar radars they were rotated many times in the ionosphere on the way to the Moon and as many times again on the way back, so that they eventually returned to the receiving aerial with a very arbitrary polarization depending upon the state of the Earth's ionosphere. Variations in the ray path through the ionosphere and in the state of ionization of the latter caused the deep fading of the Moon echoes. This effect, which was so troublesome to the early observers, was later put to good use to investigate the state of the ionosphere, although nowadays it is more convenient to use the signals from artificial satellites rather than from lunar radar.

The rapid fading was found to be attributable to the libration of the Moon, that is the way that it *apparently* wobbles on its axis as seen from the Earth. This slightly changing aspect of the Moon enables the optical astronomer to photograph 58 per cent of the surface of the Moon, for it brings into view a little of the side that is normally hidden. It affects the radar echo by causing the radar range to various minor features of the surface to change by the order of half a wavelength in a few seconds, or even less at short wavelengths. The resultant signal strength depends upon the phase of all the signals at any instant and varies very markedly over short periods.

Despite the effects of libration the Moon appears to be quite a smooth object. If parallel light shines on a ball with a slightly rough surface, like a sand-blasted metal sphere, the whole illuminated hemisphere of the ball appears to reflect back light and, although the intensity of the light falls off gradually towards the edge (limb darkening), the ball appears to be its normal size. If the same parallel light falls on a very smooth sphere, such as a ball bearing, only a tiny spot in the centre of the sphere reflects back the light. It was found with lunar radar that the reflection from the Moon much more closely approximated to this condition than it did to a roughened sphere, and so it follows that the Moon must on the average be quite a smooth object to radio waves with a surface far less

pitted than that of the Earth. Most of the energy returned comes from a disc in the centre of the Moon, about a fifth of the Moon's diameter. The behaviour can be explained on the assumption that there are surface features which are large compared to the radar wavelengths but which have gentle slopes, up to about 10°. Some of these reflect the incident radiation straight back in the direction of the receiver, but once the slope of the Moon's generally spherical surface approaches 10°, the number of facets that can reflect in this specular fashion rapidly falls off and the radiation falling on these parts is reflected into space away from the receiving aerial.

If a narrow pulse is transmitted to the Moon this will be returned approximately two and a half seconds later, very much weaker and somewhat more elongated in its duration. The leading edge of the returned pulse will correspond to a reflection from the foremost part of the Moon's surface, while the trailing edge will be formed by the contributions from the most distant parts which are able to reflect. It is thus possible to get some idea of the lunar topography by studying the way in which rings at successively increasing distances return the transmitted signals. In order to achieve less ambiguous information a telescope of high resolving power must be used and this resolving power is in excess of that which can be obtained by the conventional use of a single radar dish. It is possible to use an interferometer system for the aerial and thereby to increase the resolving power in one dimension, but a simple system of this type would be confused by the responses in the separate lobes covering the surface of the Moon. A very elegant method was proposed some years ago by Martin Ryle and has been employed on the Moon radar equipment at Jodrell Bank. This is an aperture synthesis technique (see Chapter Eight) which makes use of the relative motion of the Earth to synthesize an aerial system of very considerable length. It relies upon the fact that with modern technique it is possible to maintain the frequency of the transmitter constant and therefore the transmitted and received signals remain phase coherent during the time that the aerial moves an appreciable distance as a result of the Earth's motion. The signals can therefore be combined together in such a way as to reconstruct the effect of surveying the Moon with a knife edge aerial beam. The effect of combining this technique with accurate ranging is to survey the Moon in a series of pairs of points symmetrically distributed in the northern and southern hemispheres.

Radar studies of the planets require very much more power, sensitivity and stability than radar studies of the Moon. The power and sensitivity are required because of the drastic reduction in the signal strength resulting from the inverse fourth power of the distance in the radar equation. The effect of this is to reduce the signals from Venus to one five-millionth of those from the Moon if the transmitter is the same for both. For the other planets it is somewhat worse by further factors of 16 for Mars, 120 for Mercury and 600 for Jupiter. These factors take no account of the effect of the spin of the planets and their reflecting powers, the effect of the spin of the planet Mars, for example, would make it nearly 10,000 times harder to detect than Venus with a continuous wave (Doppler shift) radar system.

In view of the extremely weak radar echoes from the planets and the fact that the radar beam widths are so very much larger than their angular sizes, one may wonder why such an expensive technique should be used. The results of the radar measurements of Venus certainly show excellent dividends. Two important astronomical measurements have resulted from these investigations, the accurate measurement of the astronomical unit and the rate and direction of spin of the planet.

The astronomical unit is the distance from the Earth to the Sun and as this is the effective baseline for surveying, by parallax, the distance to the nearer stars, it serves as a yardstick for nearly all our measurements in the Universe. It is a curious fact that although the periods and shapes of the orbits of the planets are known with a very high degree of accuracy, the absolute sizes of the orbits are relatively most uncertain, in fact they are known to the same accuracy as the astronomical unit which, until the advent of planetary radar, was about one part in a few thousand. If the distance to any one of the planets could be found accurately in terms of miles or kilometres then the astronomical unit and the sizes of all the other orbits could be computed with almost equal accuracy. The radar studies of Venus have enabled the accuracy to be increased by one or two orders of magnitude and have, rather curiously, resulted in a value which is not within the errors of the best quoted optical measurements. The best radar measurements of the range of Venus when converted to give the number of kilometres in the astronomical unit currently result in a value of 149,598,500 $\pm$ 500 km. The best value from optical astronomical methods given by De Vaucouleurs in 1961 is 149,536,000 $\pm$ 3000 km.

In a simple earthbound radar system it is usual to restrict the repetition frequency of the pulses so that echoes from the most distant points have time to return before the following pulse is returned by nearby objects. In this way it is possible to avoid confusion in the range of the targets and, as the delay time is quite short and the echoes are strong, the technique is perfectly suitable for the purpose. With planetary radar the pulse repetition frequency required by extrapolation would be one in many minutes, even for Venus. The returned echoes would individually be very much weaker than the background noise in the receiver and the integration of sufficient pulses to obtain a satisfactory signal to noise ratio would have to last for several days. Fortunately the radar range of the planet is known beforehand from ordinary astronomical measurements to within a fraction of a second so that no confusion should arise if pulses are transmitted continuously at the rate of a few a second and if the range upon reception is 'gated' to pass and integrate pulses returned after the estimated time has elapsed. This use of *a priori* information can, however, have its pitfalls. The first radar echoes from Venus were obtained at Lincoln Laboratories in the United States in 1958. Their results agreed with the optical measurements of the astronomical unit within reasonable margins of error. In the following conjunction the same laboratories repeated the measurement and failed to obtain a significant echo. A team at Jodrell Bank also endeavoured to obtain echoes on this occasion and reported a marginally detectable echo at the range originally reported by Lincoln Laboratories; they noted, however, that there was a fair chance that this response was only a statistical fluctuation of the noise in the receiver.

A number of observatories tried again at the closest approach in 1961 and on this occasion there was general agreement between three separate teams in the U.S.A. and the team at Jodrell Bank. The value that they obtained differed significantly from the earlier range measured by Lincoln Laboratories and Jodrell Bank. On this occasion a Russian team also made the measurement, using a pulsed radar with a rather high pulse repetition frequency. They made use of the *a priori* information that the range was likely to be close to that originally quoted by Lincoln Laboratories. The result was that although they obtained a clear echo, they interpreted it as occurring in response to the wrong pulse, for they were confused by the rapid pulse repetition frequency. When the initial error of their ways be-

came apparent to them they identified the correct pulse and published an amended result which agreed with the values obtained in the U.S.A. and Britain.

If a pulsed radar is used for planetary echoes it is necessary to integrate the returned signals in order to achieve a significant signal to noise ratio. In view of the fact that the range of the planet is continually changing as a result of the orbital motions of the planet and the Earth, it is necessary accurately to compensate for the changing delay or otherwise to analyse the receiver output in such a way as to allow for the variation of range. In addition to the variation in range there is also a variation in the frequency of the received signals, for the planet has a finite velocity relative to the Earth which results in a significant Doppler shift of the signals; a component of velocity towards the Earth results in a higher frequency for the returned signals whereas a velocity away from the Earth lowers the frequency of the echoes. The receiver, therefore, must be accurately tuned to a pre-arranged programme of frequencies, or else the output must be analysed in frequency in accordance with a similar spectral law, in order that a proper integration of the signals may be achieved. The change in frequency resulting from the Doppler shift is typically a few hundred cycles per hour and with a pulsed radar system it is necessary to keep the receiving system or spectral analyser tuned to within a few cycles per second. In common with other measurements in which very faint signals are sought it is necessary to compare the output in a region where the signal is anticipated with a control region where it is unlikely to be present. In a pulsed radar system this is achieved by comparing the output at the expected range with the 'noise' output, or statistical variations in the receiver power, at ranges slightly less than and in excess of the range of the signal.

With Venus and Mercury there is an advantage in using a continuous wave transmission rather than a pulsed radar to obtain the echoes. The shift in frequency can then be measured with very great accuracy for the intrinsic bandwidth is extremely narrow. The range information is lost but in its place a very accurate value for the orbital velocity may be obtained from the precise central frequency of the returned signals. The bandwidth of the returned signals now results almost entirely from the Doppler broadening imposed by the spin of the planet, and the spin rate may be accurately assessed. With Venus this has been found to be about 300 days *retrograde*. With a

continuous wave transmission there is obviously a problem in preventing the transmitted signals from entering the sensitive receiver. It is possible to make use of the fact that a circularly polarized transmitted signal is returned upon reflection with the opposite polarization, but it is still necessary to make or break a connection in the transmitting or receiving phases. An alternative and attractive, though costly, solution is to have separate and remote large transmitting and receiving aerials; truly continuous operation can then be maintained.

With Mars, Jupiter and Saturn the rate of spin is such as to widen the likely radar echoes to such an extent that continuous wave operation may not be warranted, but it is worth noting that there are two hybrid techniques that can be applied. One of these is a modulated continuous wave signal and is tantamount to degenerate pulsed radar, the modulation used being far less deep than in a pulsed system. The second system is a coherent radar in which pulses are used and the signal is coherently detected in the receiver. This method has been used by Lincoln Laboratories and the former technique by Jet Propulsion Laboratories in their successful work on Venus.

CHAPTER FIVE

GALACTIC RADIO EMISSIONS AND THE FIRST RADIO STARS

In Chapter One our Galaxy was described as a vast spiral nebula containing about 100,000,000,000 stars embedded in a tenuous medium of gas and dust. It is a very ordinary galaxy from most points of view, somewhat smaller than its nearest spiral neighbour, the Great Nebula in Andromeda, and not at the present time undergoing a catastrophic event such as the explosions and collisions which are observed in some of the distant galaxies. By these standards it is classified as a normal galaxy of the *Sb* type, that is to say it has a spiral structure which is neither very tightly wound (*Sa* galaxies) nor very open and spidery (type *Sc*). Despite its normality there are many processes occurring in our Galaxy which serve to emit radio waves, and, from our vantage point within one of the spiral arms, the total signal strength of this radiation is considerable. It was, in fact, this radiation which constituted the very first observation in Radio Astronomy, many years before even the radiation from the Sun was detected.

In 1932 Karl Jansky, of the Bell Telephone Laboratories in America, was studying the noise level to be expected when a sensitive short wave radio receiver was connected to a directional aerial system. In 1932 the short wave spectrum was relatively free from the hordes of broadcasting stations and other services with which it is packed today and a good receiver connected to an aerial usually yielded an enhanced level of hissing noise and crackles which reduced its overall sensitivity. It was obviously advantageous to study these noises and to determine their origin so that they might be reduced by the use of a suitably designed receiver or aerial system. Jansky found that he could clearly differentiate between two types of noise, the crackling sounds that could usually be identified with thunderstorm activity and a relatively steady background hiss which was at first most puzzling.

Jansky found that the hissing noise was the strongest when his aerial was pointing in a particular direction and he could find no significant terrestrial feature with which to associate it. He kept watch for many months and then a very curious fact

emerged, the interpretation of which must stand to Jansky's eternal credit as an experimenter. The maximum of the radiation gradually changed its position relative to the Earth and the Sun but *not* relative to the stars in the sky. Jansky was also able to point his aerial in different directions and he showed that the radiation came from the Milky Way. The maximum radiation came from the constellation of Sagittarius, in the direction of the central hub of our Galaxy.

It is a remarkable fact that although Jansky's work was well publicized at the time, the significance of his observations did not appear to be appreciated by the astronomers of the day and it was left to a radio amateur, Grote Reber, brilliantly to pursue Jansky's work with home-made equipment and to show its enormous potentialities some seven years later. Reber, work- at a higher frequency with a home-made paraboloid was able to map the radiation over a large part of the sky and to show that, although it was strongest in the direction of the galactic centre, it extended with contours of varying intensity along the whole belt of the Milky Way. Reber's contour maps are a classic of Radio Astronomy and constituted the first survey of the heavens in the light of radiation remote from the visible spectrum. There was a broad correlation with the light from the Milky Way but isolated peaks of emission appeared in various regions where there were no remarkable optical features. Some of these peaks coincide with the positions of the first radio 'stars' discovered nearly ten years later in the constellations of Cassiopeia, Cygnus and Taurus.

The radiation originally detected by Karl Jansky we now believe to be generated almost entirely in the inter-stellar gas by a process known as synchrotron emission. This process was briefly mentioned in the chapter on the Sun but we must now consider it a little more closely for those who wish to understand how it occurs.

An electron moving in a magnetic field is constrained to travel with a circular motion about the field lines, so that in general its path is a helix, the radius and pitch of which depend upon the initial motion of the electron relative to the direction of the field, and its strength. The rate at which it orbits around the lines of force depends only upon the strength of the field and is known as the gyro frequency.

$$w_H = \frac{eH}{m_o c} \text{ radians per second,}$$

where $\frac{e}{m_o}$ is the ratio of charge to mass for the electron, $5{\cdot}273 \times 10^{17}$ e.s.u./g.,

H is the strength of the field in gauss, and

c is the velocity of light, 3×10^{10} cm/s.

Putting in these constants and dividing by 2π, the gyro frequency may be expressed in c/s as 2·8 Mc/s per gauss. In the Galaxy, where the magnetic field is of the order of a millionth of a gauss, free but slow moving electrons would be expected to radiate weakly at a frequency of about 2·8 c/s because they would be subject to a centrifugal acceleration as they described their helical paths and a kinematically accelerated charge radiates. This is often referred to as cyclotron radiation. Kinematic acceleration was specified to differentiate the case of a charged particle constrained in a fixed position but subject to a gravitational acceleration in that position; this charge would not radiate but, perversely, it would do so if allowed to fall freely. It is only the kinematic component of acceleration that gives rise to radiation.

The picture changes for very fast electrons. If the velocity of the electron approaches that of light we would expect relativity to play a part. From the special theory of relativity, confirmed by many experiments, we know that if a source of light or other radiation is moving at almost the speed of light as seen by an observer, O, the radiation which would appear to an observer, S, moving with the source, to be going out in a broad polar diagram, would appear to observer O to be confined to a narrow beam in the direction of relative motion. This is part of the phenomenon of aberration; the angle at which a particular ray leaves the source, in the opinion of the two observers, is different, for though they may agree about distances measured normal to the instantaneous velocity, they disagree about distances in the direction of motion as a result of the Fitzgerald-Lorentz contraction. The half angle of the cone of radiation measured by a distant observer therefore shrinks* to

* From the relativistic aberration formula $\sin\theta = \dfrac{\sin\theta' \sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v}{c}\cos\theta'}$

whence for $\theta' = 90°$, $\sin\theta = \sqrt{1 - \frac{v^2}{c^2}}$. As v tends to c, $\sin\theta$ tends to θ.

$\theta = \sqrt{1 - \frac{v^2}{c^2}}$. The observer seeing this cone may simply argue that the radiation covers a much wider angle at the source because of the Lorentz contraction in the direction of motion. A distant observer also sees the light bluer, for the radiation only reaches him when the instantaneous velocity is in his direction and a Doppler shift to the blue therefore occurs. With an electron travelling at such a speed that its energy is 3×10^9 eV, the radiation will appear to be beamed in the direction of instantaneous velocity within a cone of vertex angle 30 seconds of arc.

The second important contribution of relativity to the situation is that the mass of the fast moving electron, in the opinion of the distant observer, is not the rest mass used in the formula for the gyro frequency, but the relativistic mass $\frac{m_0}{\sqrt{1 - (v^2/c^2)}}$

To him the electron appears much heavier, so that in a given magnetic field, H, the electron appears to orbit with a slower angular frequency,

$$\omega'_H = \frac{eH}{m_0 c}\sqrt{1 - (v^2/c^2)}.$$

$\frac{\omega'_H}{2\pi}$ is sometimes called the relativistic cyclotron frequency.

The electron itself, on the other hand, experiences an angular velocity $\omega'' = \frac{\omega'_H}{\sqrt{1 - (\omega'^2_H R^2/c^2)}}$ so that if it is in a flat circular orbit of radius R it 'sees' the Universe rotating at the gyro frequency but thinks that it is describing a much tighter circle around the magnetic field lines. From either viewpoint the velocity of the electron is limited so that it cannot exceed the speed of light. In general the electrons have a component of velocity along the lines of magnetic force and their orbits will be long drawn out spirals. A distant observer will only receive radiation when the tip of the spiral is pointing closely at him (within the angle θ) and the electron is approaching him at the velocity v. He therefore receives a series of short sharp pulses spaced apart at the relativistic cyclotron frequency during the period that he is within the polar diagram or cone of radiation from the electron.

Ginzburg, Shklovsky and others have used a more mathe-

matical argument of this type but have incorporated the ratio of the rest energy of the electron to its actual energy,

$$\frac{m_0c^2}{E},$$

where $\sqrt{1 - (v^2/c^2)}$ is used in this account*. They then give the width of the individual pulses as

$$\Delta t = \frac{m_0c}{eH}\left(\frac{m_0c^2}{E}\right)^2$$

which can be written as

$$\frac{m_0c}{eH}\left(1 - \frac{v^2}{c^2}\right)$$

The spectrum from an individual electron is then the Fourier transform† of a series of these pulses occurring at a pulse repetition frequency corresponding to the relativistic cyclotron frequency. This spectrum is therefore a tightly packed series of line emissions corresponding to harmonics of ω'_H stretching up to frequencies of the order of $1/\Delta t$. If the particle energy is extremely high, θ may be so narrow that only the radiation from a single pulse will be intercepted by the observer and the spectrum will appear continuous up to a frequency of the order of $1/\Delta t$.

As stated here it may appear that the account is not entirely satisfactory from the beginning, for the electron is not a self luminous source to which concepts of Doppler shift and aberration can be applied without more detailed justification.

An electron rotating in a circle has two components of motion at right angles. It may be represented by two sinusoidal oscillations, one up and down, and the other sideways. If the circle is viewed edge on, the electron appears to be dancing up and down and the sideways motion is now towards and away from the observer. The up and down motion generates a sine wave in the direction of the observer but the fore and aft motion involves no displacement of the electric field normal to the

* The actual (or total) energy $E = T + m_0c^2$, where T is the kinetic energy; but $T = \frac{m_0c^2}{\sqrt{1 - (v^2/c^2)}} - m_0c^2$ (where the last term on the right is necessary in order that the kinetic energy shall be zero when the electron is at rest), hence $E = \frac{m_0c^2}{\sqrt{1 - (v^2/c^2)}}$ and therefore $\frac{m_0c^2}{E} = \sqrt{1 - (v^2/c^2)}$.

† See Appendix.

observer and with a *slow* moving electron it contributes little observable effect. The situation is analogous to two dipole aerials, one seen broadside and the other end-on. Only the broadside radiation is detected.

Each point on the radiated sine wave may be associated with a point of origin on the electron orbit, thus the crests correspond to the top of the orbit, the troughs to the bottom and the mean zero level to the points where centre of the orbit, the electron and the observer are in line. Now any simple sine wave may be sampled in a series of separate intervals which are each much shorter than the period of the wave. If these intervals follow one another with no gaps, the samples may be recombined to give a reproduction of the original wave, but any single sample will appear as a short pulse and will have a spectrum containing many high frequencies quite unlike the original wave. Each

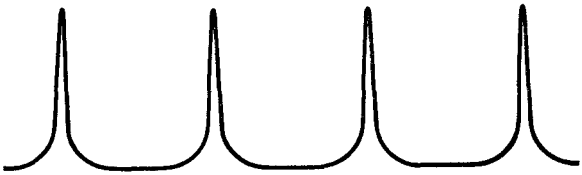


Fig. 14. The waveform from a single electron orbiting at a high speed in a magnetic field. The peaks become sharper if the speed of the electron approaches closer to that of light. In most astronomical contexts only a single pulse may be received from each electron

sample corresponds to the signal transmitted as the electron moves through one of the equally divided sampling intervals in the plane of the orbit. Now consider what happens to the wave generated by an electron describing a circular orbit at a very high speed. At the top of the orbit the electron is moving towards the observer and the wavelengths in the little sample will be Doppler shifted to the blue, the frequency components of the sample will therefore be higher than in a sine wave. At the bottom of the orbit the electron is receding and the frequencies in the sample are shifted to the red, the wavelengths will be long and drawn out (they may not be seen at all for a *very* fast electron). Upon recombining all the samples the waveform will not now be that of a sine wave but a waveform containing sharp peaks (Fig. 14). The frequency spectrum of this waveform extends to very high frequencies corresponding to the rise time of the narrow peaks, as might be

guessed from the blue shift occurring at these instants. At velocities very close to that of light the effect of the flashing narrow cone of radiation predominates, and one must also take into account the slower apparent rate of rotation caused by the relativistic increase of mass.

In practice there will be not one electron but myriads of them performing a variety of helical orbits according to their energies and angles of launch relative to the direction of the magnetic field. All these electrons will contribute to the ultimate synchrotron radiation from the region but these contributions will usually be incoherent. Not only will their initial phases differ but the fundamental frequency corresponding to the relativistic cyclotron frequency is dependent upon the individual particle energy as well as the precise field strength at the particle. The fact that the spectrum is related to the particle energy has an interesting consequence. It means that if it can be established that the radiation from a certain region of space is synchrotron radiation, then the spectrum of cosmic ray particle energies responsible for the radiation may be determined.

In order to obtain synchrotron radiation we have seen that it is necessary to have the combination of relativistic electrons and a magnetic field. The Galaxy certainly has a magnetic field of about a millionth of a gauss and it is likely that it is permeated by cosmic ray electrons, some of which are generated in a variety of stellar eruptions and explosions and some of which may be accelerated on a grander scale within the Galaxy itself. From recent work originally pursued by Charles Seeger it has been found that much of the radiation from the Galaxy is linearly polarized and that this polarization follows a reasonably well ordered pattern. Thus the theory can account for the general background of radiation which is observed predominantly at lower frequencies from the Galaxy and it also points to the likely mechanism for the intense generation of radiation in the localized sources that have come to be known as radio stars.

The term 'radio star' is something of a misnomer; it was originally applied to the few intense sources of emission that had been observed up to about the year 1951, for all attempts at determining their angular diameter had met with no success and, to the aerial systems of the day, they appeared to be star-like points in the sky. The first of these radio stars was discovered by that wonderful man J. S. Hey, who had previously discovered the solar radio emissions by his wartime detective

work on radar interference. After the war Hey set out to survey the sky in the manner of Jansky and Reber with an aerial system of high directivity. He obtained a detailed map showing the contours of the radio Galaxy but one curious feature caught his eye. When the aerial was directed towards the constellation of Cygnus the signal strength was not only enhanced, it was fluctuating. Hey's keen mind correctly interpreted this as a discrete source of radiation which was either varying its intrinsic power output or scintillating in a similar way to the appearance of visible stars. There was no obvious feature in the direction of Cygnus from which this curious radiation could come. The first of a long line of wonderful new objects had been discovered. We now know that the fluctuations that Hey observed were not intrinsic in the source but were produced by patches of varying ionization scurrying about in the Earth's upper atmosphere. The correlation of this twinkling at points up to a few miles apart on the Earth's surface and the measurement of time delays between the scintillations at these points, gave rise to a very useful technique for studying the winds in the F layers of the Earth's ionosphere.

John Bolton and Gordon Stanley, in Australia, successfully recorded the fringes produced by the Cygnus source with an interferometer on the cliffs near Sydney. The constellation of Cygnus rises only just above the horizon in Sydney and Bolton was able to receive the signal on a single aerial system by two paths, one direct and the other by reflection from the sea. The relative path distances varied as the source moved across the sky and caused the signals to occur alternately in and out of phase at the aerial in a manner analogous to the interference obtained from a Lloyd's mirror in optics. Not only did Bolton confirm Hey's localized source in Cygnus but another small intense source came to light in the constellation of Taurus and this, alone among the early radio stars, was quickly identified. Its position coincided with that of a well known curious object, the Crab Nebula.

The Crab Nebula is shown in Plate 8. It appears as a glowing skein of filamentary excited gas clouds with an angular diameter about a quarter of that of the Sun. It appears to be expanding and from the rate of expansion it is possible to extrapolate backwards to the time when it could have started from a point. This fixes the origin of the nebula as some catastrophe which occurred in the heavens in the year 1054. Now in the year 1054 there were not so many astronomers but there were quite a few

astrologers who depended for their living, and their lives, on a constant study of the heavens. In particular the Chinese astrologers at this time kept detailed records of their observations and the messages that they read into them. The Chinese astrologers chronicled one such observation thus:

> 2nd cyclical day, 5th month, 1st year of Chih-ho of Sung, guest star appearing at South East of T'ien-kuan, several ts'un long, lasting more than one year.

A modern reporter would have recorded that in the year 1054 a faint star in the constellation of Taurus suddenly blazed forth as bright as the planets and then gradually faded once more into obscurity. The Chinese astrologers had witnessed a supernova explosion, a rare event in the Galaxy and one that many present day astronomers would forego all other observations to witness. Only two others have been reliably recorded in our Galaxy, by Tycho Brahe in 1572 and by Kepler in 1604. Tycho Brahe was a brilliant observer and his documentation of the event is extremely accurate despite the fact that he had no telescope. Kepler was the man who showed from Tycho Brahe's observations of the planets that they moved in elliptical orbits and, though he was primarily a theoretician and inferior to Tycho Brahe as an observer, his measurements were nevertheless of a high professional standard. In both cases it was recorded that a star flared up by many magnitudes and then faded away until after a year or two it was hardly observable. Unlike the Crab Nebula these supernovae have not left remnants which are brilliant landmarks, though they have both been detected as radio sources, many times weaker than the Crab. Present-day observations with large optical telescopes of supernovae explosions in other galaxies show that statistically there should have been many more in our own Galaxy than have been recorded. The reason for this discrepancy is that most of them occur near the galactic plane in regions where the light is obscured by the dust in the spiral arms, a supernova near the galactic centre would be quite invisible from the Earth. The remnants from such invisible explosions contain a plentiful supply of high energy particles which radiate in a weak magnetic field to contribute to the features of the radio sky.

To witness the birth of a supernova is virtually to witness the death of a star. An ordinary nova explosion is rather like a star sneezing, it casts off its outer mantle and flares up in so doing but may repeat this process many times during its life. With

a supernova there is just one gigantic explosion involving the whole body of the star and it is surprising that any stellar body should remain after such a catastrophe. In the Crab Nebula there does appear to be a small parent star embedded in the centre of the expanding remnants but all the power and the glory are in the remnants themselves. The precise case history of a star which leads to the ultimate attainment of the status of a supernova is not fully understood but the ultimate explosion is thought to be a culmination of a process of nuclear fusion that makes man's efforts puny in the extreme. Most stars depend for their energy on a controlled thermonuclear reaction whereby they convert hydrogen into helium. This process can continue for hundreds of millions of years with little change in the external properties of the star but all the time a larger and larger core of helium is accumulating in the stellar interior. Eventually a stage is reached when nearly all the hydrogen has been consumed, the interior of the star is intensely hot and the helium nuclei combine to form the heavy elements. This process, once started, is most rapid and the energy released rends the star asunder.

The brightest of all the radio stars in the sky, Cassiopeia A, is thought to be the shell of a supernova that flared up nearly three hundred years ago but escaped optical detection. Cassiopeia A was first observed by Martin Ryle in 1947. He had constructed an interferometer aerial system to investigate the Cygnus A source reported by Hey and not only received Cygnus and Taurus but also an even more powerful signal from the constellation of Cassiopeia. Intensive efforts were made during the next few years to pinpoint these sources, to endeavour to resolve their angular diameters and, with Cygnus and Cassiopeia, to identify them with optical features. Up to the early 1950s all efforts to resolve or identify the sources failed; they appeared to have angular diameters less than 6 minutes of arc, the highest resolving power then in use, and no features in the sky appeared at all unusual within the broad regions within which they could be pinpointed. In the year 1950 it seemed to many astronomers that they might really be dark and mysterious stellar objects.

Three teams participated in the next phase of the attack. At Cambridge F. G. Smith accurately measured the position of the sources and enabled D. W. Dewhirst to photograph a faint but curious nebulosity near Cassiopeia. F. G. Smith followed up these positional measurements with measurements of the visibility (see Chapter Eight) of the sources in the East–West

direction and was able to detect a slight reduction in their visibility at his longest baselines. Meanwhile B. Mills in Australia, using a much longer East–West baseline, clearly resolved Cygnus but could not see Cassiopeia with his aerial. At Manchester the author and M. K. Das Gupta had constructed a new type of interferometer which had been devised by R. Hanbury Brown. It later became known as the Intensity Interferometer and it was capable of a very high resolving power. With it they measured the two sources in a North–South direction and, clearly resolving them in this direction, immediately followed up with measurements of the shape by using different orientations of their interferometer. Cassiopeia turned out to be almost circular and about 4 minutes of arc in diameter. Cygnus at first appeared anomalous, the readings just would not fit together on the assumption that the source was a disc, an ellipse or a generally irregular distribution. The author then pointed out that everything made sense if Cygnus was, in fact, a binary source with two centres of emission spaced apart by about one and a half minutes of arc in Right Ascension. Shortly afterwards Jennison and Das Gupta were able to show from a series of measurements along the long axis of the source that this was indeed so.

The curious feature photographed by Dewhirst was sufficient incentive for Baade and Minkowski of the Mount Wilson and Palomar Observatories to use the 200 in. Palomar Telescope to survey F. G. Smith's positions. Close to the position of Cygnus there was a cluster of galaxies containing a very curious nebula which they interpreted as two galaxies in collision. This interpretation is now questioned, but as it appears that Cygnus A is an extra-galactic object the discussion will be deferred to a later chapter. In the position of Cassiopeia, Baade and Minkowski photographed a very faint but remarkable emission nebulosity occupying a disc a little over 4 minutes of arc in diameter with its most northerly edge most clearly defined and coinciding with the part that Dewhirst first detected. Over the remainder of the disc the nebulosity was only apparent upon close inspection of the plates for it consisted of a few short and disconnected filaments and condensations of emission. Nevertheless Baade and Minkowski were able to obtain spectra of 3 of the more diffuse filaments and these showed a remarkably high degree of excitation. Cassiopeia A appeared to be a highly energetic supernova shell of a type which had not previously been identified in the Galaxy.

In the middle 1950s the Cassiopeia source was intensively studied by the author at radio wavelengths and by Baade and Minkowski at optical wavelengths. The author developed a new interferometer technique in which the signals from 3 widely spaced aerials were brought together by radio links and compared in such a way that the phase of the signals as well as their amplitudes could be accurately measured. Together with V. Latham he was able to show that Cassiopeia exhibited pronounced limb brightening, consistent with the radiation from a shell rather than the volume of the nebulosity, and it had a faint spur stretching out to one side to a distance of 6 minutes of arc from the centre of the nebula, the main body of which had a diameter of 4·1 minutes of arc. Rowson and others showed that the diameter of the main body was very similar at much higher frequencies. At Mount Palomar, Baade and Minkowski took a number of photographs over a period of a few years and showed that the sharp condensations appeared to be stationary as though the expanding shell had run into gas clouds in the interstellar medium which had substantially reduced the outward velocity. These sharp condensations only emitted red light, mainly two of the spectral lines H_α from hydrogen and $[N_{II}]$ from nitrogen. The diffuse filaments were much bluer and emitted a great many spectral lines of high excitation which showed Doppler shifts corresponding to velocities as high as 5000 km/s. These diffuse filaments changed their shape in a very short space of time and were almost unrecognizable over a period of years. From the motion of these filaments Baade and Minkowski were able to show that the nebula was about 280 years old (Fig. 15) while their measurements and independent hydrogen line absorption measurements of the radio astronomers gave the distance as nearly 10,000 light years. The description of the Cassiopeia source presented by Minkowski at the Paris Symposium on Radio Astronomy in 1958 showed a remarkable agreement with the radio picture synthesized by the author and V. Latham from interferometer measurements at 125 and 132 Mc/s. The angular diameter of the main body of the nebula agreed almost exactly and Minkowski had also detected the spur on his photographic plates as a very faint extension which was barely observable on one side only of the nebula. This spur is a curious feature which showed up clearly and unambiguously with the author's phase-sensitive measuring technique but the existence of it was contested by a French group who had made measurements on a higher frequency

without this facility. It is possible that the spur could be a transient feature and that it could have a spectral characteristic which strongly favoured the lower frequencies. The clear records of Jennison and Latham combined with Minkowski's independent optical observation certainly point to the reality of its existence but give no clue as to how it could be formed. If it is a jet of gas which started from the centre of the nebula 280 years ago the most distant parts would have to be moving away at over 15,000 km/s, the faint appearance of the jet is

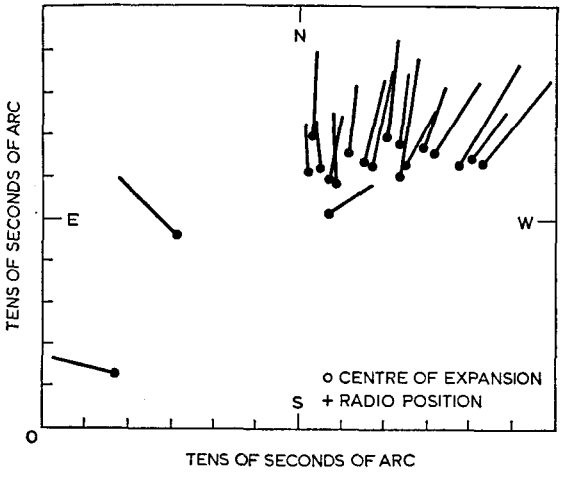


Fig. 15. The motion of individual diffuse filaments in the Cassiopeia Nebulosity. The outgoing lines represent the observed proper motions extrapolated over 100 years (after Minkowski)

generally diffuse and it may well be that the excitation is other than that from a stream of gas.

Apart from Cassiopeia A, the Crab nebula in Taurus and Tycho Brahe's and Kepler's supernovae there are at least two objects in the Galaxy which are radio sources and have every appearance of being the remains of very old supernovae. One of these is IC 443 in the constellation of Gemini, a beautiful object consisting of large and luminous clouds with sharply defined edges which could well be parts of the shell of a supernova of bygone days. A rather similar object is the Cygnus loop nebula, a great pair of tenuous brackets in the sky which has every appearance of the shell of a supernova.

Galactic radio sources of the supernova type are very transient on a cosmic scale but, although the radiation from Cassiopeia A is definitely getting weaker, they are not usually categorized as variable sources. There is another type of radio star in the Galaxy which is highly variable and these objects are actually real stars. They are a category of red dwarf stars that have come to be known as flare stars for they occasionally emit great flares of light which considerably increase their luminosity. These flares are on a much more gigantic scale than the familiar solar flares and Unsold and others predicted long ago that there was a possibility that they might emit detectable radio signals. The confirmation of this fact was provided after much patient observation by A. C. B. Lovell using the 250 ft Jodrell Bank telescope and correlating his records with the observations of visual observatories. In particular he collaborated with F. W. Whipple who arranged for a few selected stars to be photographed by the satellite tracking Schmidt camera network in the periods between satellite passes. The continuous manning of these cameras combined with their facilities for accurate timing rendered them highly suitable for catching the elusive flares. On the very first occasion that Lovell pointed his telescope at a flare star he recorded a sudden burst of intensity which at the time could not with certainty be identified with radiation from the flare star, for transient interfering signals occasionally gave rise to not dissimilar bursts. It was some years before he was able to obtain unambiguous records which showed good correlation with the optical observations and none of these quite rivalled the first recording which, from the behaviour of the trace, was in retrospect almost certainly genuine. The radio signals (Fig. 16) usually arrived within a few minutes of the optical outburst but occasionally they appeared to be almost simultaneous and on at least one occasion the radio outburst may have preceded the optical burst, although this could possibly be accounted for by the discontinuous optical record from the Schmidt cameras. Besides the obvious astronomical and radio astronomical interests of this work, it provided an excellent demonstration of the equality of the velocity of propagation of radio and light waves over great distances. From the correlation of the optical and radio data the two agreed to within the order of one minute in 10 years.

All the galactic radio emissions which have been discussed in this chapter may be categorized as non-thermal emissions.

Fig. 16.
(a) and (b) Radio outbursts correlated with
(c) optical flare on flare star UV Ceti

The radiation is much stronger than that which could originate from tenuous clouds radiating only by virtue of their natural thermal temperatures, and the spectra of the sources are incompatible with the laws of thermal radiation. There is another class of radio sources which radiates entirely by a thermal process, the radiation is in general much weaker than that from the non-thermal sources but it may be detected in emission above about 20 Mc/s and in absorption, dark against the bright non-thermal background, at lower frequencies.

This class of sources embraces H II regions of our Galaxy. H II regions are clouds of inter-stellar hydrogen which are excited to a temperature of about 10,000°K by their close association with hot stars of type O and B. The hydrogen gas absorbs the utra violet radiations from these stars and is excited so that it radiates a dull red light which may be detected photographically. Perhaps the best known H II region is the great nebula in Orion, for this can be seen with the naked eye forming the star-like object at the top of Orion's sword. A small telescope or pair of binoculars reveals this object as a nebula, and a large telescope shows that it contains a group of very hot stars. Other well known H II regions are the Rosette nebula, which contains four hot stars, and the Trifid nebula, which is rather more distant and appears more compact than that in Orion.

The radio frequency emission from H II regions simply corresponds to their temperature as hot clouds of gas and it cannot therefore exceed the actual temperature of the excited gas. This temperature is usually limited at 10,000°K by the intervention of other radiative processes, in particular the radiation from atomic oxygen. The temperature recorded by a radio telescope will depend upon the optical depth* of the cloud at the frequency of measurement, so that the radio frequency spectra of H II regions show a change of slope at a frequency corresponding to that at which the whole nebula is optically thick; this frequency is usually of the order of 100 Mc/s. The concept of optical depth is discussed in Chapter Six, in which will be found an account of the emission from the H I regions or neutral hydrogen gas clouds of the Galaxy.

* See footnote on p. 64.

CHAPTER SIX

SPECTRAL MEASUREMENTS

The information that radio astronomers can collect to enable them to reconstruct details of the radio sky and to solve the mysteries that their pictures unfold, is primarily obtained from measurements of intensity and direction but is tinted here and there with a little colour. The angular resolving power of a radio telescope is very poor compared to that of an optical telescope but the spectral resolving power, which determines the ability to separate different wavelengths of the radiation, corresponds to the selectivity of the radio receiver, and this may be very high. To use this high selectivity for most purposes would be unwise for the noise ripple on the output of the receiver increases as the selectivity is increased and it is preferable for such purposes as spatial surveys to widen the band width as far as possible in order to gain the square root of this factor in the ultimate sensitivity. This process would smear out any narrow spectral features in the radiation from the Galaxy and radio stars but it may be justified by the fact that no such features *associated with the sources* have yet been found, though there are a few spectral lines associated with the inter-stellar medium. The sources appear to have very broad spectra, commensurate with synchrotron radiation or thermal emission, according to the class of the source, and though they may fade over 10 octaves, the radiation appears to contain no sharp peaks or troughs of emission which could be used, for example, to determine the red shift of the most distant sources.

Nevertheless the very broad spectra of the radio sources are worthy of some attention to see if they can be categorized in any way from their slopes, and in the hope that this categorization may correlate with other facts that we may know about the sources. Spectral measurements of solar phenomena have already been discussed. These measurements fall into a different category, apart from the applicability of a similar treatment to flare stars; solar spectroscopy will not be discussed in this chapter.

The determination of the spectrum of a radio source requires the measurement of the flux density (the power received per unit band width on a unit area at the Earth) in absolute terms

at several, and preferably all, frequencies. Absolute measurements have been made at relatively few frequencies, principally 38, 400, 1400 and 3200 Mc/s. The usual technique is to sample the power received from the source in an aerial having an accurately known effective area, and to compare this with the noise power generated in a heated resistor. Relative intensity measurements may be made at other frequencies using a 'standard' source, such as Cassiopeia A, for comparison. Cassiopeia A is gradually fading so that it is unwise to use it for relative comparisons over long periods, it also has the disadvantage of a large angular diameter and a much higher intensity than most of the sources with which it may be compared. Conway, Kellerman and Long in a concerted attack on the spectra of the sources used 7 substandards for comparison, the sources 3C123, 161, 196, 348, 353, 380, and 409. They listed the sources in 5 classes according to the appearance of their spectra on a log/log plot.

Class	*Spectrum*
T	Thermal sources
S	Straight slope over limited observed frequency range
S_1	Straight slope from 38 to 3200 Mc/s
S_2	Straight slope to about 1400 Mc/s then more rapid fall at higher frequencies
C	Curved spectrum, i.e. the spectral index is a function of frequency

They estimate that between 30 and 50 per cent of all sources have spectra that are straight between 38 and 3200 Mc/s. The straight line for any source could be expressed by the relation

$$\log S = x \log f + \text{constant}$$

where x is the spectral index.

Those sources which fall into categories S_2 and C all deviated in the sense that the spectrum became more convex, the spectral index becoming more negative with increasing frequency. More recent measurements have shown that a few sources have spectra which are concave over part of the radio frequency spectrum.

A few of the sources whose spectra have been measured have been identified with photographic objects. Four supernova remnants are all classed as S_1. Three sources believed to be Quasars, 3C 48, 3C 196 and 3C 286, are of spectral class

C, S_2 and C respectively. Cygnus A is of spectral class C. This information probably just tells us that the same emission mechanism is at work in these sources and that this is very likely the synchrotron mechanism. The spectra of those radio sources whose distance is known differ so much that it is not at present possible to use this information to determine the red shift of an individual radio source. There appears to be no correlation of the spectral indices with the distance of identified sources, nor with their absolute luminosity. There *does* appear to be a correlation between spectral curvature and brightness temperature; the sources of high brightness temperature tend to be of spectral class C.

THE HYDROGEN LINE

The hydrogen line was briefly mentioned in discussing the basic tools of the trade. It is a spectral line from neutral hydrogen on a wavelength of 21 cm (1420·4 Mc/s) which was predicted as a possible component of cosmic radiation by H. C. Van de Hulst in 1945. Van de Hulst was working at Leiden during the German occupation of Holland, on a problem in theoretical astronomy suggested by Professor Oort. He investigated the possibility of the formation of hitherto undetected radio frequency spectral lines in the Galaxy, and as hydrogen is by far the most abundant element he looked very carefully into the possible low energy transitions of the hydrogen atom. He predicted that if a hydrogen atom was in its unexcited or ground state, designated by $^2S_{\frac{1}{2}}$, the magnetic fields of the proton and electron could change from parallel to anti-parallel orientations and a spectral line should be emitted at a frequency of 1420 Mc/s. There appeared to be a good chance of detecting this transition from the vast but sparsely distributed neutral hydrogen content of the Galaxy.

It was five years later before technology was sufficiently well advanced to enable the line to be detected, and, as usual on these occasions, there was a close race to the finishing post which was won this time by Ewen and Purcell in the U.S.A. The Dutch team had remarkably bad luck on this occasion for their receiver was destroyed by fire shortly before they were ready to set up their experiment.

The strength of the signal emitted from a large cloud of neutral hydrogen depends upon its temperature and the total number of atoms in the line of sight. If there is a very large number of atoms the cloud is opaque and the signal strength

corresponds to the temperature of the cloud provided that the cloud fills the aerial beam. If the cloud has a small angular diameter the signal strength falls in the ratio of the solid angle subtended by the cloud to the solid angle of the aerial beam. The brightness temperature, T_b, from the cloud itself, that is the temperature that would be recorded if it filled the aerial beam, is given by*

$$T_b = T(1 - e^{-\Delta}),$$

where T is the 'temperature' of the gas corresponding to the population of the two quantum states (parallel and anti-parallel),

and Δ is the total optical thickness of the gas.

If the gas is thin and therefore almost transparent, the optical thickness is small, Δ is much less than unity, and

$$T_b = T\Delta.$$

If the gas is thick and opaque, Δ is greater than unity and the cloud is equivalent to a black body at 21 cm so that $T_b = T$. A hydrogen line receiver in these circumstances would record a temperature of about 100°K from the clouds of hydrogen in the Galaxy.

If the story ended here the hydrogen line would only contribute to our knowledge of astronomy by enabling us to plot the temperature or opacity of the hydrogen clouds in various directions. This would certainly be very useful information, for nearly all the other radio signals come from the regions where the hydrogen atoms are ionized, but there is another property of the hydrogen line, a property of very far-reaching importance, which has not yet been mentioned. The frequency f' at

* Adding all the elementary contributions along a ray, the brightness temperature in the direction of the ray is given by

$$T_b = \int_{-\infty}^{\infty} e^{-\tau} T \, d\tau,$$

where T is the temperature of the gas and τ is the optical depth to each element. For a thick layer of isothermal homogeneous gas of temperature T and optical thickness Δ,

$$T_b = \int_0^{\Delta} e^{-\tau} T \, d\tau$$
$$= T(1 - e^{-\Delta}).$$

T_b approaches T exponentially as the thickness of the layer increases.

which each atom emits radiation is 1420·405 Mc/s but the frequency, f, at which this radiation is received, depends upon the velocity of the emitting atom relative to the observer according to the Doppler effect,

$$f = f' \frac{1 + (v/c) \cos \theta'}{\sqrt{1 - (v^2/c^2)}},$$

where θ' is the angle of motion at the source relative to the observer. For typical velocities of galactic hydrogen clouds in the direction of the line of sight, this expression may be simplified to

$$\frac{\delta f}{f'} = \frac{v}{c},$$

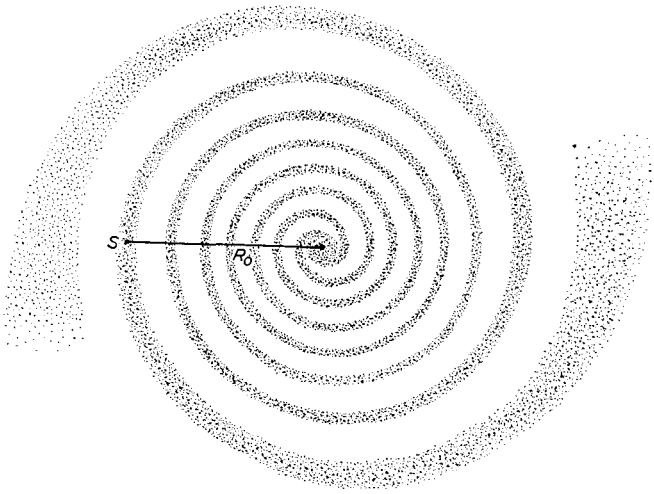
where $\delta f = f - f'$ is the frequency shift produced by the velocity v and is in the sense of a shift to lower frequencies (longer and redder wavelengths) for receding velocities. Inserting the values for the frequency of emission of the hydrogen line and the velocity of light it will be seen that a velocity of approach or recession of 1 km/s produces a Doppler shift of 4·73 kc/s. The number of atoms radiating at a given observed frequency is the number moving with the corresponding radial velocity. Of course the atoms do not have to move directly towards or away from the observer, it is primarily the *component* of velocity towards or along the line of sight that dominates the Doppler shift, and this can be seen from the factor Cos θ' in the equation connecting f and f'. If a sensitive receiver is constructed in the form of a spectrometer so that the strength and frequency of the received signals can be measured, then the component of motion of the emitting atoms in the line of sight may be established. In general there will be random motions of individual atoms resulting from their finite temperature and quasi-random motions resulting from turbulence within the neutral gas clouds comprising the so-called H I regions of the Galaxy. These random motions are equally likely to produce positive and negative components of velocity so that their effect is to broaden the spectral width of the line. Turbulent velocities greatly exceed the thermal velocities so that only the former contribute appreciably to the broadening of the line.

The H I regions take part in the general rotation of the Galaxy and, in addition, individual gas clouds have random velocities within the spiral arms. The clouds themselves are

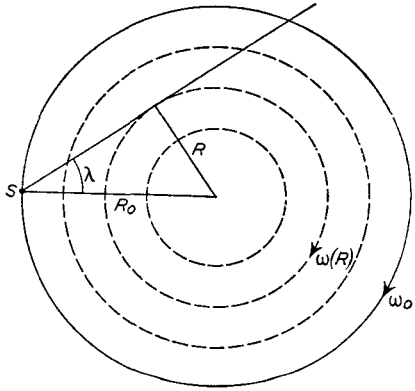
typically 5 or 10 parsecs across (about 20 light years)* so that the nearby clouds completely fill the beam of a modern radio telescope, whilst clouds in the furthermost parts are usually much smaller than the beam width. If there is only a single cloud in the aerial beam the hydrogen line emission from it will be received at a frequency centred upon the Doppler shift appropriate to the component of motion of the cloud towards or away from the Earth and it will be broadened as a result of the random velocities in the cloud. The component of motion in the direction of the Earth will be composite. It depends upon the velocity of motion of the Earth around the Sun, the direction of motion of the Earth at the time, the velocity of the Sun in the galactic system and the velocity of the cloud in the galactic system. The velocity of the radio telescope resulting from the fact that the Earth is spinning on its axis is a mere 1000 miles/h and can usually be neglected compared to the 20 miles/s velocity of the Earth in its orbit and the much larger velocities of many parts of the Galaxy.

The fact that the Galaxy rotates was recognized before the discovery of the hydrogen line but observations of the velocity distribution within the galactic plane were rendered extremely difficult by the obscuration of the inter-stellar dust. Nevertheless Oort and others had produced tentative models for the way in which the Galaxy rotated and, according to these models, it appeared that it was not rotating in the manner of a rigid spiral with the same angular velocity applicable at all parts but rather in the manner of the spring of a clock in which the turns are slipping relative to one another. With these models as the basis for analysis, the early hydrogen line measurements showed that it was possible to determine the structure of the Galaxy in remarkable detail by measuring the Doppler velocities in various directions in the galactic plane and interpreting these as the differential velocities corresponding to the slipping of the spiral arms as the whole system rotates. The radio studies were unaffected by the dust clouds so that even the remote regions of the Galaxy could be observed. From the way in which the observations as a whole fitted together it was

* The parsec is the unit most frequently used by astronomers for the measurement of large distances. One parsec is the distance at which an object would have to be placed in order that the radius of the Earth's orbit would appear from that object to subtend one second of arc. One parsec equals 3·258 light years = $3{\cdot}08 \times 10^{13}$ km = 206,265 astronomical units. The abbreviation kpc represents a unit of 1000 parsecs.



(a)



(b)

Fig. 17. Galactic rotation
(a) idealized spiral nebula
(b) line of sight at an angle λ to the line joining the Sun to the galactic centre in the circular approximation

possible to revise the original rotational models so that the worst anomalies in the original analysis could be removed, and as a result of this process of feedback our knowledge of galactic structure has made gigantic strides in recent years.

The solar system is located in a spiral arm of the Galaxy at a distance, R_0, of about 8 kiloparsecs (kpc) from the centre. Within the circle described at this radius from the centre and upon the assumption that the rotation may be represented by a series of rings each rotating rigidly but each slipping relative to its neighbours, the radial velocity relative to the solar system of any point distance R from the centre and at an angle λ to R_0 at the Sun is given by

$$v_r = R_0[\omega(R) - \omega_0]\sin\lambda,$$

where $\omega(R)$ is the angular velocity about the centre of the ring at radius R,

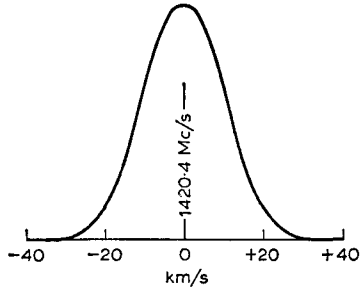
ω_0 is the angular velocity of the Sun about the centre.

It will be seen from Fig. 17 that for a given direction λ (less than 90°) the observed velocity v_r goes through a maximum, where the radius vector from the galactic centre is perpendicular to the line of sight, and then diminishes until it is zero again at the second intersection with the ring at radius R_0, thereafter for more distant regions, the sign changes, as it also does for observations at negative values of λ, on the opposite side of R_0. The distance from the centre to the position of maximum velocity at a particular angle λ is given simply by the right angle triangle in the figure as $R_0 \sin\lambda$. Spectra near $\lambda = 180°$ and $\lambda = 90°$ are shown in Fig. 18.

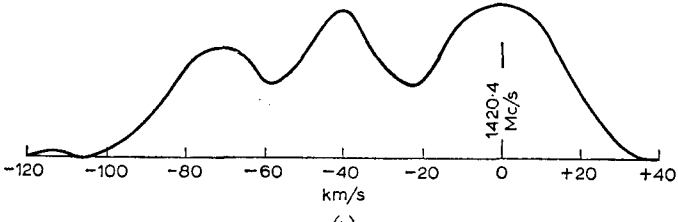
$\omega(R)$, the variation of angular velocity with radius, has been determined with the aid of these relationships by Kwee Müller and Westerhout from the northern hemisphere of the Earth and by Kerr from the southern hemisphere. The two sets of measurements map different parts of the Galaxy but their rough agreement for regions between 3 and 8 kpc from the centre reasonably justifies the model with circular rotation in this range of radii. In the inner region within 3 kpc of the centre the situation is complicated by the Galaxy itself, for there appears to be an expanding arm and other curious features, and by the inaccuracy introduced through the value of sin λ becoming small.

Analysis of the hydrogen line spectra from H I regions in the vicinity of the Sun is especially difficult because of the small radial velocities and the increased sensitivity of the analysis to

the effects of local random velocities of the clouds. Current maps of the H I distribution in the Galaxy show a general structure in the form of a number of near-circular arms but there is a curious convergence of the arms in the neighbourhood of the Sun. The chance of the Sun being located at a nodal point of this kind is extremely unlikely and it is much more likely that the analysis is still imperfect. There is general agreement



(a)



(b)

Fig. 18. Hydrogen line emission spectra
(a) λ (Fig. 17) $\approx 180°$
(b) $\lambda \approx 90°$

on the thickness of the neutral hydrogen layer in the solar neighbourhood and on the density of the gas in this region, the thickness is about 200 parsecs and the density is approximately half an atom per cubic centimetre. The position of the Sun relative to the galactic plane is more uncertain, some investigators place it about 50 parsecs above the plane while others believe it to be in, and some slightly below, the plane.

The situation within 3 kpc of the galactic centre has been studied by the Dutch astronomers, Rougoor and Oort. Between the Sun and the centre of the Galaxy is a spiral arm, the so-called 3 kpc arm, about 10 light years from the centre of the Galaxy. This arm rotates with a tangential velocity of about 200 km/s but it also appears to be expanding outwards from the centre with a velocity of 50 km/s. The thickness of the arm (in galactic latitude, that is, at right angles to the plane of the Galaxy) is 120 parsecs, a little over half the thickness in the vicinity of the Earth. Rougoor and Oort find that the gas on the far side of the galactic centre appears to be streaming out even faster than that on the near side, perhaps at velocities up to 200 km/s. They compute that all the gas should be removed from the central region in between 10 million and 100 million years if there is no replenishment. This time scale is much less than the age of the Galaxy so that gas must stream in from some other source in order to maintain the system if the expansion theory is correct.

The picture of the conditions in the central parts of the Galaxy that Rougoor and Oort developed from their observations is as follows. There appears to be a disc of neutral hydrogen at the centre and in the innermost regions the density of the gas must be very high, 1000 atoms/cc or more at a distance of 10 parsecs, falling off rapidly with increasing distance until there is practically no gas at a little over 300 parsecs. At a distance of 500 parsecs the neutral hydrogen appears again, as a ring 100 parsecs wide and 80 parsecs thick, rotating at about 265 km/s. The inner disc also appears to be about 80 parsecs thick and is apparently rotating at about 200 km/s near its edge. Neither the disc nor the ring appears to be expanding in the manner of the 3 kpc arm.

Up to now we have been considering the emission of the hydrogen line, the way in which the frequency of 1420·4 Mc/s is radiated from different parts of the Galaxy. The H I regions which emit the hydrogen line may also be detected in absorption if, behind the neutral hydrogen gas, there is a bright source emitting a broad or 'white' spectrum of radiation. In optics the same phenomenon is seen in the spectrum of the Sun and many of the stars, dark lines, such as the Fraunhofer lines, are seen silhouetted against the bright white light of the Sun and they originate in upper, cooler layers of the Sun's atmosphere. The hydrogen line in absorption is seen against the strong continuous radiation from some of the radio stars. A cool dense

cloud between the source and the Earth absorbs the radiation in a narrow range of frequencies corresponding to the Doppler shifted hydrogen line spectrum of the cloud. Usually there is a number of clouds between the source and the Earth and these clouds belong to different spiral arms so that many absorption lines are seen. In some cases the line of sight from the Earth to the radio star may only cut the edge of a cloud and only very partial absorption is then likely to occur. If the radio source subtends a large angular diameter at the Earth, a cloud may only obscure part of the source so that again the absorption is only partial unless the beam width of the telescope is capable of resolving the smaller feature.

The ease with which one may detect absorption lines depends upon the optical depth of the obscuring cloud and the intensity of the radio source, if it is assumed that the cloud covers the whole source. For a cloud which is optically thick, it is evident that all the radiation from the distant radio star will be absorbed in the cloud at the frequency which to the cloud is 1420·405 Mc/s and which to us is this frequency appropriately Doppler shifted according to the velocity of the cloud relative to the Earth. If a narrow band spectrometer is tuned through this frequency, the signal strength to either side of the line will depend upon the brightness of the radio star, but the signal strength will dip down to register the temperature of the hydrogen when it is exactly tuned to the Doppler shifted line. In these circumstances, when the aerial temperature from the radio star considerably exceeds 100°K, the signal of the hydrogen line in absorption is comparable to that of the radio star observed with the same narrow bandwidth. The bandwidth for this purpose should be less than the spectral width of a single hydrogen cloud, which is typically about one kilocycle.

Absorption measurements provide a technique for studying individual hydrogen clouds, for the angular diameter of the radio stars is typically a few minutes of arc so that only a narrow pencil of rays joins the source to the Earth and the few clouds through which this pencil cuts can usually be separated in depth by their differing Doppler shifts. The Doppler shifts of these clouds are usually similar to those of other clouds in the same spiral arms so that it is possible to measure the distance to certain radio stars by studying the absorption profile of the source. A classic example of the use of this method of measuring distances occurred with the Cassiopeia source. From the motion of the filaments of this source the optical observers

originally claimed that it was only about half as far away as the distance determined by radio astronomers who had observed an absorption feature which they identified as belonging to a more distant spiral arm. The optical astronomers argued that this feature was probably a cloud in a nearer spiral arm and that it had acquired a higher random velocity than was typical for most clouds. The radio astronomers remained adamant and in the end the optical astronomers discovered an error in their own analysis, which, when removed, gave good agreement with the radio result.

EXTRA-GALACTIC AND INTER-GALACTIC SPECTROSCOPY

The success of the hydrogen line absorption measurements for the estimation of distances to galactic sources encouraged efforts to look for similar features in the extra-galactic sources, for this would at last enable the radio astronomers to estimate the distance of some of the objects which are now of considerable cosmological interest. The first measurements were made by Lilley and McClain; they compared the spectra of Cygnus A with that of Cassiopeia A at frequencies close to 1340 Mc/s, which would correspond to the cosmologically red shifted hydrogen line if it were present in or near Cygnus A. These results were unfortunately confused by instrumental difficulties. A long series of more sensitive measurements by the author and R. D. Davies, with a special scanning spectrometer that they had constructed for the purpose and which they used in conjunction with the 250 ft Jodrell Bank telescope, failed to show any significant features in either emission or absorption in the neighbourhood of 1340 Mc/s. It appears that there is insufficient neutral hydrogen in the vicinity of Cygnus A to absorb the continuous emission. Any hydrogen which has escaped the catastrophe which has overtaken this galaxy may well be excited by the radiation density in the neighbourhood. A faint absorption feature has been found more recently with the Australian 210 ft dish in the spectrum of the extra-galactic source in Virgo. This source is considerably nearer than Cygnus and the measurement is unfortunately not typical of the cosmological red shift from the distant sources. The Australian dish is more efficient than the Jodrell telescope at 21 cm and the absorption feature measured by the Australians was marginally below the threshold of the measurements made on Virgo in Britain.

Field in the United States and R. D. Davies and the author

in Britain, have searched for inter-galactic neutral hydrogen. If neutral hydrogen is present between the galaxies it might be detected in emission from all parts of the sky or in absorption where it intercepts the rays from a distant radio source. In emission the spectrum that would be expected would come

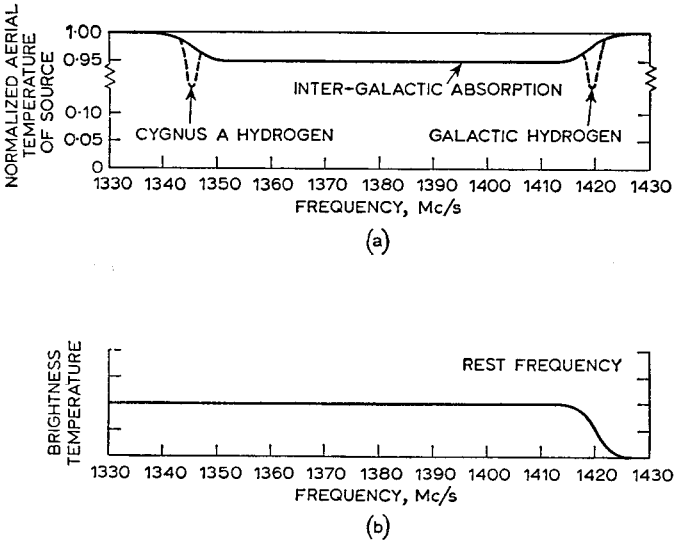


Fig. 19. The effect (not yet observed) of inter-galactic hydrogen
(a) in absorption against Cygnus A
(b) in emission from all parts of the sky

from neutral atoms throughout the whole universe. The emission from these atoms would be red shifted according to their distance from our Galaxy and the spectrum would extend all the way from the lowest frequencies up to 1420 Mc/s (Fig. 19(b)). In absorption against the Cygnus A source, the spectrum that would be expected from inter-galactic neutral hydrogen would result from all the atoms present in the intervening space absorbing at their appropriate red shifted wavelengths. The spectrum would therefore be a trough, perhaps broken by the presence of discrete clouds, stretching from 1340 Mc/s (the frequency of the hydrogen line with the red shift appropriate to Cygnus A) to the frequency of the local standard, 1420·4 Mc/s, and crowned at the high frequency end by absorption in the spiral arms of our own Galaxy (Fig. 19(a)).

The most sensitive of these measurements, by Davies and the author, showed no detectable step in emission nor trough in absorption. They put the limit on the optical depth in absorption as 0·0008. It is very likely that any matter between the galaxies is not present in the form of cool neutral hydrogen.

The hydrogen line in emission has been detected from a number of normal nearby galaxies. The first to be detected were the Magellanic clouds, the two smaller irregular companions to our own Galaxy. Later the giant nebula in Andromeda, M 31, was detected and efforts are being made to survey its structure in detail. From the observations of the emission of the hydrogen line from a number of galaxies of different types it is possible to compute the ratio of the mass of each galaxy to the mass of neutral hydrogen that it contains. On the observations to date it appears that the irregular galaxies such as the Magellanic clouds contain proportionately more hydrogen than the spirals. Some of the measurements are given in the following table. The table is arranged in descending order of the proportion of neutral hydrogen.

TABLE 2

Galaxy	*Type*	*Ratio of neutral hydrogen mass to galactic mass*
IC 1613	Irr	0·15
Small Magellanic cloud	Irr	0·125
NGC 6822	Irr	0·10
Large Magellanic cloud	Irr	0·08
M 33	Sc+	0·06
M 82	Irr	0·055
M 101	Sc−	0·02
Galaxy	Sb?	0·02
M 31 (Andromeda)	Sb−	0·01
M 81	Sb−	0·01

It is unfortunate that it is unlikely that detection of the emission of the hydrogen line from normal galaxies will add much to our knowledge of the distant parts of the Universe and the cosmological problems that become manifest at such distances, but the measurements do provide information on the galaxies themselves and how they may evolve individually and as a family.

ZEEMAN SPLITTING

The hydrogen line has a property which it shares with other spectral lines and which may help us to understand much more of the conditions in the Galaxy. The hydrogen line exhibits Zeeman splitting in the presence of a magnetic field. If the magnetic field is directed along the line of sight, the line is split into left- and right-hand circularly polarized components separated by 2·8 Mc/s per gauss. If the field is directed away from the observer and the line is seen in emission, the left-hand component is displaced slightly above 1420 Mc/s and the right-hand component is displaced by the same amount to lower frequencies. The displacement is reversed if the field is reversed or if the line is seen in absorption. Absorption measurements provide the most sensitive situation, for the effective signal strength of an individual line is enhanced in silhouette against the background and the individual clouds which are so revealed are likely to be permeated by relatively homogeneous magnetic fields so that the displaced lines retain a narrow profile.

The strength of the magnetic fields that have been determined to date by R. D. Davies and his colleagues at Jodrell Bank, lie in the range from a millionth of a gauss to about 25 times this value. The measurements have been made on the galactic absorption lines of Cassiopeia A, Taurus A, Cygnus A and Sagittarius A and a few isolated clouds in emission. The results when taken together are consistent with a general magnetic field along the spiral arms with a strength of a few millionths of a gauss.

The technique used to measure the Zeeman splitting on the hydrogen line is worth brief comment. As the line splitting is proportional to the magnetic field, it will be seen that in the magnetic field of the Galaxy the lines will only be separated by about 10 c/s. As the width of the spectral feature from an individual cloud may be many 100 or 1000 c/s it would not be possible to isolate the two peaks simply by using the resolving power of a narrow band spectrometer. The method of measurement that is used in practice makes use of the opposite sense of circular polarization of the two components to separate them in the receiver. The radio telescope is equipped with a circularly polarized aerial which is continuously switched between the two senses of polarization whilst the frequency of the receiver is swept across the absorption feature. The output of the receiver in the two switched states is continuously compared by

synchronously reversing the output circuit prior to integrating. If the line is split by the Zeeman effect, the strength of the signal in one polarization exceeds that in the other on one side of the line, whilst the position is reversed on the other side, so that the output of the receiver swings from negative to positive as it is tuned through the central frequency. On a pen recorder the output looks like a dollar sign on its side, the line through the $ representing the undeviated output of the receiver.

OTHER SPECTRAL LINES

Attempts have been made to detect other spectral lines from the Galaxy. In particular the line from deuterium at about 300 Mc/s has been searched for on many occasions but with no definite success. Recently a line from the OH radical was observed by Weinreb, Barrett, Meeks and Henry in the absorption spectrum of Cassiopeia A. The line is double with rest frequencies of 1665·402 and 1667·357 Mc/s and there are satellite lines at 1612·2 and 1720·5 Mc/s which have been observed by the Australians in the absorption spectrum of Sagittarius A. These new lines will provide a great deal of further information about the conditions in the Galaxy and, in particular, the distribution and abundance of OH. It is probably unwise to speculate at this stage but it is unlikely that they will surplant the hydrogen line as the major tool of the radio spectroscopists.

CHAPTER SEVEN

THE EXTRA-GALACTIC RADIO SOURCES

There are two radio stars which shine far brighter than all the others. One of them is the Cassiopeia A radio source which we now know to be a supernova remnant in our own Galaxy, but the other, two thirds as bright, and perhaps one day the brightest of all, is a fascinating and curious object so far away in the depths of space that the radio waves and the light from it have taken over 500,000,000 years to reach the Earth. This remarkable radio source was the very first to be discovered, the radio source first observed by Hey in the constellation of Cygnus. In the years after Hey's discovery it defied all attempts to identify it with any conspicuous visible object until, in 1952, it finally yielded to Baade and Minkowski at the 200 in. telescope at Mount Palomar.

The object which Baade and Minkowski identified as the Cygnus A source was a peculiar, distorted nebula amongst a loose cluster of galaxies. It measured approximately 18 times 30 seconds of arc and had two small bright nuclei near the centre, spaced apart by only two seconds of arc. The object appeared to be two galaxies in collision, tidally distorted by their gravitational fields. There is a story that Baade first proposed this interpretation to Minkowski who challenged him to prove it by showing that its spectrum was consistent with the violent excitation that would be expected in such a catastrophe. The spectrum was duly obtained and clearly showed from the high excitation of the emission lines that the nebula was indeed in a state of agitation consistent with a collision. Baade won the bet, and with it a bottle of whisky. The observed spectral lines and their identifications are tabulated on page 78.

The brackets around all but one of the spectral lines indicate that they are so-called 'forbidden' lines corresponding to transitions which are not readily observed in the laboratory. The last column lists the recessional velocity, upon the assumption that the red shift of the observed lines is a Doppler shift, computed for each individual line with the mean, 16830 km/s given at the bottom. Using the current value for the Hubble constant, relating the velocity to the distance of the nebula, the distance of Cygnus A is found to be 170,000,000 parsecs or about 550,000,000 light years.

The red shift places most of the spectral lines in a region where the sensitivity of the original photographic plates rendered the assessment of their intensities inaccurate, but some features were clearly distinguishable. More than 50 per cent of the total light is contained in the emission lines while the lines themselves are severely distorted. The Hα emission is relatively faint in comparison with the adjacent [N II] lines, while Hβ is not seen despite the fact that the neighbouring [O III] lines are strong. The [O II] line is very strong and extends over the whole visible size of the nebula in position angle 90°. The high

TABLE 3. EMISSION LINES IN CYGNUS A

Observed line, Å	*Identification*	$\frac{-cd\lambda}{\lambda}$ km/s
3819·9	3428·8 [Ne V]	17010
3937·2	3726·06/8·82 [O II]	16930
4087·5	3868·77 [Ne III]	16930
4189·6	3967·48 [Ne III]	16820
5234·3	4958·91 [O III]	16660
5284·6	5001·85 [O III]	16955
6642·5	6300·27 [O I]	17770
6718·6	6363·88 [O I]	16720
6916·6	6548·06 [Ne II]	16870
6928·0	6562·66 Hα	16300
6949·2	6583·43 [N II]	16670
		Mean 16830

excitation forbidden line [Ne V] is readily discernable and is indicative of a highly energetic process at work in the nebula. All the lines are diffuse but the widths are difficult to measure. The velocity spread is of the order of 400 km/s.

The catastrophe which Baade and Minkowski believed to have overtaken Cygnus was a wonderful phenomenon and it led to similar interpretations for a number of other radio sources during the following decade. The galaxies, in addition to apparently receding statistically from one another in accordance with the usual interpretation of the red shift, are also milling about with their own superimposed random velocities which may be of the order of 1000 km/s. They are packed in intergalactic space with a density more or less in proportion to that

of the air liners over an international airfield. What could be more likely than that they would occasionally collide. Indeed, Zwicky had obtained some wonderful photographs of galaxies connected by trails of matter which clearly showed the signs of strong interaction commensurate with a bygone collision. Within the time scale of the Universe the statistical probability of the collision of galaxies in a cluster is very high and, in the collision, the energy released would be expected to be phenomenal. This release of energy was expected to occur through a collision of the gas clouds in the two systems, for most of the stars would shoot past one another without physical contact. This may seem surprising until one remembers that, unlike the relatively high spatial density of galaxies in inter-galactic space, the spatial density of the stars in inter-stellar space resembles model aeroplanes a few thousand miles apart.

The precise mechanism by which the collision of the gas clouds could give rise to the intense radio frequency radiation was not fully explained, but Shklovsky suggested that as they were locked into magnetic fields they would bounce off one another and would be traversed by magneto-hydrodynamic shock waves which would result in the release of energy by a synchrotron process. Certainly the radio frequency spectrum of Cygnus A appeared to be consistent with a synchrotron mechanism and there appeared to be sufficient energy available in the collision to provide the radiation over the million years or so that a collision would be expected to last.

For nearly a decade after the identification of Cygnus and its interpretation as a pair of galaxies in collision, the majority of astronomers and astro-physicists were convinced that galactic collisions were responsible for all, or nearly all, the other extra-galactic radio sources. At an international symposium in 1955 a very famous astronomer, who had been closely associated with this work, delivered a paper on the identification of the radio sources from which it is enlightening to quote:

> The identification of a source due to a collision between galaxies presents relatively little difficulty. Large nearby systems of this kind can easily be recognized from their appearance. Actual collisions seem to lead to conspicuous emission of forbidden lines of high excitation which can still be observed in objects too distant and thus too small to reveal their nature from their appearance. While it seems not yet possible to predict the strength of the radio emission

from the appearance of a given object, the sequence of conditions seems basically clear. Close pairs of galaxies such as NGC 4575/4576, without visible signs of interaction, do not show enhanced radio emission. Close pairs with large tidal interactions, such as NGC 4038/4039, may give somewhat enhanced radio emission. . . .

Strong interaction is represented by systems like Cygnus A, Hercules A, and NGC 1275. The most powerful source of this type, Cygnus A, is unfortunately too distant, thus too small and faint to permit a detailed optical investigation. . . .

NGC 1275 is the only known sample of colliding galaxies which is suitable for a detailed optical investigation. That this object and not the Perseus cluster of galaxies is to be identified with the small source in its position is now safely established. The system consists of a tightly wound spiral of early-type and a strongly distorted late-type spiral. Inspection of blue and red exposures shows some unusual features regarding the colours and distribution of emission patches connected with the distorted spiral arms of the late-type system. . . . In the northern part of the object, two sets of lines appear, separated by about 3000 km/s, which show unmistakably the presence of two separate gas masses. Comparison of the structural details on the direct photographs with the spectrograms shows that the velocities of the early spiral and of the late-type system are about +5200 km/s and +8200 km/s respectively. Since the absorption patches of the late-type systems are obviously in front, the northern part of the late-type system moves towards the early spiral. . . .

The interpretation of these results is obvious. The two galaxies, which are seen nearly face on, are in collision. They are inclined towards each other in such a way that originally the southern part of the late system was closer to the early spiral. As the galaxies moved towards each other, the collision started in the South and progressed to the North, where interaction between the separate gas masses of the two galaxies is now going on. Farther to the South, probably from some line slightly to the North of the nucleus, the collision is over, leaving the combined gas mass formed by the collision in a highly heated and excited condition. The velocity of the combined gas differs little from that of the early spiral; this requires that the mass of the gas in this system be considerably higher than that in the late-type galaxy. This

result does not appear improbable. The early spiral is a system of very high luminosity having an absolute magnitude $M_{pg} \approx -19$ and may be expected to contain a correspondingly large mass of gas. More detailed discussions of the conditions in this system have to wait for additional observations.

If one uses a crude model in which the gas masses are considered as plane parallel layers of 500 parsecs thickness, one finds that the observed conditions require an angle of 15° to 20° between the galactic planes of the two systems. On this basis, the total duration of the collision is of the order 10^6 years. The intensity of the radio emission during this period should first increase, reach one or possibly several maxima depending on the distributions of gas in the two galaxies, and finally decline. If the present power represents about the average, the total energy emitted during the collision must be of the order of 10^{47} erg, still a small fraction of the total kinetic energy of 10^{49} to 10^{50} erg available in a collision between galaxies with a relative velocity of 3000 km/s.

It is a remarkable fact that, after such strong, and apparently almost conclusive, evidence had been obtained for the association of extra-galactic radio sources with galactic collisions, the concept was later shown to be apparently unworkable and it was completely dropped by the the early 1960s. Before considering how this came to pass we must see what information the radio astronomers were able to provide. Cygnus A is the classic example of a radio galaxy; what does it look like to the radio astronomer?

In 1951 the author was engaged with M. K. Das Gupta at Jodrell Bank in a race against his colleagues F. G. Smith in Cambridge and B. Y. Mills in Australia to measure the diameter of Cygnus A. As most of the equipment of its day, the Jodrell Bank interferometer was home-made. It was a new concept in technique that had been suggested by Hanbury Brown. It had a very high resolving power and operated over radio links but it required very large aerials. It is now known as the post-detector or intensity interferometer and it is described in Chapter Eight. The aerials used with this interferometer were not like the big dishes for which Jodrell Bank is now famous, they were tiltable wooden structures 120 × 40 ft, supporting collinear-broadside arrays of many hundreds of dipoles and

designed to collapse like folding tables so that they could be transported about the countryside where and when the local farmers permitted. The first measurements on Cygnus were made with one of these aerials at Jodrell Bank and the other in the paddock of Dr. (now Sir Bernard) Lovell's house, 4 km South of Jodrell Bank.

Jennison and Das Gupta lost the race to *resolve* Cygnus by a short head, but they were fortunate enough to discover and follow up one of those fascinating little anomalies that make research worthwhile. Their measurements did not give the same diameter for Cygnus as did those of Smith and Mills, and, although the technique that they used was a new concept in interferometry, they were reasonably sure of their measurements. Smith and Mills were both good observers, but they had used different baselines and measured Cygnus East–West.

The cumbersome array of dipoles was uprooted from the paddock and set up in turn at the same distance from Jodrell (or as near to 4 km as the cows and crops would permit) in two directions making angles of 120° with the first, North–South baseline. The three measurements made thus with identical equipment, showed that Cygnus was highly elongated in a roughly East–West direction, but the measurements were still anomalous. The measurements in the three orientations at the same distance should just have shown the projection of the source into the three different directions, yet the readings did not fit together. If the readings were correct, then the law of projections could not be a simple trigonometrical relationship but had to vary with azimuth in a very curious way.

At this point the author recalls that he considered the problem in his bath and left it an hour or two later shivering but jubilant in the idea that it all fitted together if Cygnus were two blobs. (A binary source would change the laws of projection in accordance with the fact that the measured parameter, the Fourier transform of the brightness distribution across the source, is not just changed in scale but is a radically different function along the major and minor axes. These functions are approximately cosinusoidal and Gaussian respectively. The former varies more rapidly with the spacing of the aerials than the latter, even for comparable source dimensions. As the baseline is rotated in azimuth a gradual transition occurs between the two functions, giving rise to the effect that was observed.) In the months that followed, Jennison and Das Gupta picked their way between the smallholdings with the trans-

portable aerial system and showed that the measured visibility of Cygnus along the new-found major axis was indeed that of a source with two centres of emission spaced about one and a half minutes of arc apart. In order to establish the result quickly they used for the first time a variation of the interferometer technique, determining the gradient of the visibility function by changing the wavelength and not the physical separation of the aerials. Not only did this speed up the operation but it considerably reduced the number of irate farmers around Jodrell Bank.

The binary structure of the Cygnus A source is shown on Plate 9. It consists of two almost symmetrical blobs, each about one minute of arc long in the direction of the major axis and about thirty seconds of arc wide in the perpendicular direction. The centres of emission are spaced apart by about 1·5 minutes of arc near position angle 100° at 127 Mc/s. The whole object is very much larger than the nebula seen by Baade and Minkowski; it completely straddles the nebula so that the maximum radio emission comes from regions where no light is seen, far away to the sides of the photographic object.

The binary interpretation of Cygnus was not universally accepted for many years but the objections were mainly on theoretical grounds concerning the legitimacy of the analysis on a minimum number of observations and with a technique which was not phase-sensitive. At that time no interferometer working on such long baselines (long by the standards of those days) could record the phase of the visibility function. The intensity interferometer measured the square of the visibility function so that it was inherently incapable of measuring the phase. A knowledge of the phase is necessary if one is to define uniquely the structure of the source, and it was highly desirable that it should therefore be obtained for Cygnus, for it was just possible that the second maximum of the visibility function along the major axis might have been in the same phase as the first maximum and not reversed as it would be for a nearly symmetrical binary source. In addition, phase measurements would be able to show whether or not the two blobs were precisely equal in intensity and, if not, which side was the brighter.

It was primarily for this purpose that the author devised the three station or phase-sensitive interferometer that he and Latham used successfully to measure the structure of Cassiopeia A. It was constructed in 1954 and when used to measure

the structure of Cygnus A it clearly showed that the source was indeed double, that the two emitting regions were remarkably matched in their brightness for such widely separated objects (the radio source is over 100,000 parsecs across) and that there was less than 10 per cent of this brightness in the central region.

The Jodrell Bank measurements were all made at frequencies between 115 and 135 Mc/s. Rowson repeated the measurement, without phase information, at 3000 Mc/s and showed that the structure hardly changed. The centres of emission appeared a little farther apart (1′ 40″) and the major axis very slightly rotated (108°). Much more recent measurements at 900 Mc/s, 1420 Mc/s, 3000 Mc/s and 10,000 Mc/s have confirmed that there appear to be slight changes in the structure at higher frequencies. Lequex finds that the two component sources are brightest near to their outermost edges. It has been suggested that there is a central component, or bridge, with a different spectral index which would give more radiation from the central region at 130 Mc/s than the limit determined by Jennison and Latham. This is based upon extrapolation from very much higher frequencies and detailed observations at low frequencies must be made before one can be sure of these slight variations.

In 1961, C. H. Mayer of the U.S. Naval Research Laboratory, after conducting a series of very carefully controlled experiments, showed that the Cygnus A source exhibited a small amount (7·5 per cent) of linear polarization at 3 cm wavelength but that the degree of polarization fell to 1·5 per cent at 5 cm and no polarization was detectable at 10 cm. Though many radio astronomers expected polarized radio waves to be generated in the Cygnus source, it is very unlikely that, at that time, any would have dared to predict that the polarization would be observable, even at 3 cm wavelength. Linear polarization is consistent with synchrotron radiation and therefore with the presence of relativistic electrons and a magnetic field, but the magnetic field must be very well ordered in order that a resultant polarization may be detected. The Cygnus source extends over about half a million light years and the observations of linear polarized emission would imply that the magnetic field is approximately coherent over an object of this size. A magnetic field and an ionized gas are strongly coupled by the conductivity of the gas, and it appears that the catastrophe that has overtaken Cygnus A has created and maintained a

remarkable regularity over truly enormous distances in the distribution of the magnetic field and the associated structure of the radio emissive regions. Linear polarization will not be detected if the emission is from regions where the magnetic field has a large variety of directions within the source where the radiation actually occurs, and it may also be rendered inobservable by differential Faraday rotation of the polarized waves as they pass through intervening ionized gas on their way to the observer. This is probably why the polarization is not detected at longer wavelengths. With Cygnus, Mayer found that the effect of Faraday rotation was exceedingly large, about 1200 rads/m^2 compared to 16 rad/m^2 for the Taurus A (galactic) source and 52 rad/m^2 for Centaurus A. The position angle of the electric vector extrapolated to zero wavelength was found to be 148° for Centaurus A, which is 79° from the line joining the two components of the source, and 33° for Cygnus A, which is 76° from the line joining the two components of the radio source. Mayer, Hollinger and Allen also found that there was no detectable circular polarization from Cygnus A, Centaurus A or Virgo A.

Mayer's work on the polarization of Cygnus and other sources, including Centaurus A, was the result of long, difficult and patient study and it deserves greater recognition than it has received. It was somewhat over-shadowed by the greater publicity later awarded to one of the first successes of the 210 ft telescope at Parkes in Australia when Bracewell, on a short visit from America, changed the polarization of the telescope as it scanned through Centaurus A and showed that the polarization varied over the source.

Centaurus A is one of the brightest radio sources in the southern hemisphere (Plate 10). It was identified, as a result of Bolton's early work, as the peculiar nebula NGC 5128. NGC 5128 appears as a nearly spherical ball with a dark twisted band running right across it, rather like the layer of dust which divides a spiral galaxy, such as in the nebula which is seen edge-on in Plate 1. On the other hand NGC 5128 does not appear to be a spiral galaxy and elliptical galaxies should have little or no gas and dust. Baade and Minkowski decided that it was a combination of an early-type system with its major axis in position angle 45° and a late-type system containing much gas and dust in position angle 135°. The object has a G-type absorption spectrum with strong emission lines, and Baade and Minkowski claimed that this spectrum showed internal

motions which supported their view that two galaxies were in interaction. More recent investigations by Burbidge and by Evans have shown that the differences of velocity between the rotating gaseous component in and near the dust lane and the stellar component are far too small to be consistent with a collision. Burbidge suggests that it may be an elliptical nebula surrounded by a ring of dust which is rotating rather more rapidly than the galaxy and is also falling in. He points out, however, that this model is unsatisfactory from a dynamical point of view.

The polarization of the Centaurus source has already been mentioned, and from the *radio* viewpoint there is also another remarkable similarity to Cygnus A. Some years after Cygnus A was found to be double, it was found that Centaurus was also double, in fact it is a double-double. There is a broad double source nearly ten degrees across merging with the galactic structure in that region, and in the centre of this gigantic double source there is a binary nucleus in which the components are separated by about 5 minutes of arc. The whole object extends for about 2,000,000 light years and yet shows a semblance of order in its magnetic fields.

The galaxy NGC 1275 has already been mentioned; it was originally thought to be the collision of two spiral nebulae, one an open Sc-type spiral and the other a tightly wound spiral of type Sa. The Sc spiral was thought to show considerable tidal distortion and the two galaxies that were thought to be present in the optical object appeared to show a relative velocity of about 3000 km/s. NGC 1275 appears as a single object at radio frequencies unlike Cygnus A and Centaurus A.

A bright radio source that did not fit well into the category of colliding galaxies was the source Virgo A. This was identified with the bright but peculiar galaxy M87 (NGC 4486) (Plate 11). The optical object appears to be an elliptical galaxy of type E_0, about 5 minutes of arc in diameter. The radio source is similar in size. The peculiarity of the optical object is a blue 'jet' which emerges from near the centre and protrudes well outside the main body of the nebula. Shklovsky suggested that this was indicative of synchrotron emission, and, sure enough, it transpired that the blue light from the jet was polarized, consistent with the generation of the light by a synchrotron process.

The downfall of the collisional hypothesis for the radio sources has already been mentioned. It came about because of two major difficulties. One of them was the problem of energy

and efficient energy conversion. In Cygnus A a total energy of 10^{60} to 10^{61} ergs is required, and if this is to be produced by the collision at 1000 km/s of two galaxies each containing a mass of about 10^{10} Suns, then the conversion of kinetic energy to radio energy must involve a highly efficient process. This argument alone is not sufficient to condemn the idea for there is probably just sufficient energy available and the masses and velocities involved are not known accurately but only assumed. The second objection is very difficult to overcome. Many of the collisions appear to involve elliptical galaxies, and elliptical galaxies contain very little gas. How then is the kinetic energy dissipated and whence comes the synchrotron radiation? No satisfactory answer has been found to this question and despite the fact that we know that galactic collisions should occur, the collisional hypothesis for the radio sources has had to give way to other ideas.

Fred Hoyle's fertile mind first came out with the idea that the energy arises from the rotational energy of a single galaxy. The magnetic field in the central region was to be wound up by the rotation and the energy finally released as a galactic flare. This theory was short lived. There is insufficient rotational energy available in most elliptical systems, and, as with the collisional hypothesis, there is the problem of the small amount of available gas. Hoyle later suggested that the ultimate source of energy was gravitational and this energy is, of course, available if it can be converted into radiation.

Shklovsky envisaged a process in which supernovae occurred in the central regions of a galaxy at such a rapid rate that there was a continuous flux of particles with the energy ultimately accountable to the nuclear energy in the exploding stars. G. Burbidge also considered that the source was nuclear energy from supernovae but postulated a chain reaction of these events among the tightly packed stars near the centre of an elliptical galaxy. He considered that the detonation wave in this process might trigger nuclear chain reactions in the interiors of up to 10,000,000 stars. Burbidge also pointed out that if, in order to assess a lower limit for the total energy of a Cygnus-like source, one equated the magnetic and particle energy density, the magnetic fields turned out to be of the order of thirty millionths of a gauss. He considered this value to be unreasonably large and suggested that the magnetic field was much smaller, with the result that it would be unable to restrain the particles and the bulk of them would escape from the magnetic field

before being able to radiate away much of their energy by the synchrotron process. The lifetime of the radio source would therefore be very transient by cosmical standards and would be unlikely to exceed 1,000,000 years. In this connection it should be pointed out that it is not known if the radiation from many galactic and extra-galactic sources is a surface emissive or volume emissive phenomenon. It will be recalled that the galactic source Cassiopeia A was found to be limb brightened and it could well be a surface emissive object, radiating from a relatively thin shell.

Since about 1960 the number of identified radio sources has risen very rapidly as a result of the accurate positions determined by the radio astronomers principally at Malvern, Cambridge, Sydney and the California Institute of Technology. The number of sources for which diameters and approximate structures have been determined, has also increased considerably, primarily through the work of H. P. Palmer's group at Jodrell Bank and Moffet and Maltby at the Owens Valley observatory of the California Institute of Technology. Palmer has used a radio link interferometer with the aerials 84 miles apart forming a baseline 500,000 wavelengths in length, and has found that a large proportion of the small, and presumably distant, radio sources are double, but that others appear to be single condensations, some with a bright central peak. Some of the objects have exceedingly small diameters, less than a fraction of a second of arc, and here unfolds another story.

In 1960, Sandage found, in the position of one of these small diameter sources, not a galaxy of any of the types that we have discussed so far, but an object which he identified as a star. The radio source was 3C 48 and the optical object was of 16th magnitude with a small area of very faint nebulosity adjacent to it. The spectrum of the stellar object showed broad emission lines which were not at first identified, and it was originally thought that the object was a star in our own Galaxy. In 1963 Greenstein and Mathews showed that the emission lines could be identified if they had been subject to a large red shift ($\delta\lambda/\lambda = 0{\cdot}36$). Meanwhile Hazard, using the 210 ft radio telescope at Parkes in Australia, determined the position of another of the small sources, 3C 273, with an accuracy of one second of arc. This was achieved by using the principle of lunar occultation. The source was in a favourable position for the Moon to pass in front of it, thereby cutting off the radiation for a short period prior to the reappearance of the source at the opposite

limb. By accurately timing the moments of disappearance and reappearance it is possible to ascertain the position of a source with exceptional precision. Hazard was not only able to determine the position of the source, he also found that it was a binary consisting of two emitting regions each less than one second of arc separated by 20 seconds of arc. The two components have different spectral indices, and it is now believed that one of them is itself a binary. An examination of the positions on photographic plates showed that one component of the source coincided exactly with a star of 12th magnitude while the second component was near the end of a luminous wisp which was directed away from the star. Schmidt at Mount Palomar found that the star had an excess of ultra violet radiation and that its spectral lines could be identified if a 20 per cent red shift correction were applied.

These two sources, 3C 48 and 3C 273, were the first of a new category of cosmic objects which have the following general properties:

(i) they look like stars,
(ii) they are usually associated with radio sources,
(iii) they have an excess ultra violet output,
(iv) they show large red shifts.

The quasi-stellar appearance of these objects has given rise to the name 'quasar' and, after the characteristics of the first few had been recognized, many more similar objects were discovered. By the end of 1964 details of 35 quasars had been published, though the red shifts had only been measured for 4 of them. One of these objects is the source 3C 286 (Plate 12), which has been studied by Shklovsky who concludes that the change in wavelength may be as much as 86 per cent. This figure is based upon the identification of a single spectral line and must therefore be accepted as very tentative. The optical spectrum of the source 3C 147, on the other hand, contains a number of lines which can be recognized and, if the red shift is assumed to be a Doppler shift corresponding to a cosmological expansion, the source must be at a distance of 5,300,000,000 light years which would make it the most distant object for which reliable measurements are available at the time of writing.* Sandage has pointed out, however, that if, in support of Shklovsky's claim, the source 3C 286 is plotted on a graph of

**Note added in proof.* The source 3C 9 has now been found to have a much greater red shift ($d\lambda/\lambda > 2$).

optical brightness against red shift, it would be close to a line determined by the other identified quasars and it may well be that the 86 per cent red shift is correct. This source has not yet been resolved at radio wavelengths; it has an angular diameter less than 0·4 second of arc and is presumed to be single unless or until further measurements prove it otherwise.

A remarkable property of the quasars is that their light output appears to be variable. The object 3C 273 can be found on photographs in the Harvard Observatory files going back over 70 years. There appear to be roughly periodic variations of about 40 per cent in its brightness with a time scale of the order of 10 to 20 years. Recent measurements show that the source also varies at radio frequencies and H. P. Palmer and his group, using an interferometer with a spacing of half a million wavelengths at 1420 Mc/s between Jodrell Bank and Malvern, have shown that the variations are associated with the smaller of the component sources. Similar but less detailed optical data is available for some of the other quasars. The source 3C 2 has been identified with a quasar and this was photographed with the 48 in. Schmidt telescope at Mount Palomar in 1951 and 1962. On these photographs it was faint and red. In 1964 it was again photographed and it was found that it had turned very blue and trebled its luminosity.

The short periods in which the output can change so drastically must mean that the majority of the radiation comes from a region which can only be of the order of, or less than, a few light years across. If the objects are really at distances of thousands of millions of light years then enormous energies are radiated from these exceedingly small volumes. Hoyle has suggested that this energy comes from the gravitational collapse of enormous clouds of gas. Fowler has proposed a model in which a gas cloud with a mass of about a million times that of the sun contracts into a region a light year across before equilibrium is reached between the inward gravitational pull and the outward pressure of radiation. He claims that this core might pulsate every few years, releasing about 10^{55} ergs of gravitational energy in each pulse. Gold has suggested that the cloud would condense into separate stars which would collide as the collapse proceeded and thence produce the flashes. The original gas cloud will in all likelihood have some angular momentum and, unless a process is envisaged which can shed this, the collapsing cloud may break up into discrete eddies to form stars or flatten into a disc-like structure.

The assumption has been made that the quasars are very distant objects, for their red shifts, if interpreted according to Hubble's law, would place them at distances of typically 5 to 10 thousand million light years. It is possible that the objects are much closer and are receding at velocities appropriate to the red shifts. Hoyle and Burbidge have recently suggested that the quasars may be objects shot out from relatively nearby galaxies in a gigantic firework display. The absence of observed blue shifts, corresponding to approaching bodies, may present a difficulty with this hypothesis.

It can be argued that the red shift could be of a different type. For example both gravitation and rotation can give a red shift and, though there would be difficulty in accounting for the observations with a rotational hypothesis, the gravitational case is at least interesting. Einstein's General Theory of Relativity predicted that a massive body should cause a shift to the red of the radiation which escapes from it. This was looked for and to some extent confirmed in the spectrum of the radiation from the Sun. It has recently been measured to an accuracy of 5 per cent by Brault and Dicke who selected the sodium D_1 line for the measurement. Further laboratory confirmation of the gravitational red shift has been provided by experiments using the Mössbauer effect. Is it possible that the quasars are nearby stars which have condensed to the extent where the gravitational field can shift the wavelength of the light by whole octaves to the red? It is known from the theoretical work of Schwarzchild in 1916 that a sufficiently massive object can not only shift the light to the red but also, in principle, make it completely disappear, trapping the radiation within a radius which is finite at the star but invisible to the distant observer. This effect is a result of the 'curvature' of the space around a gravitating body but a simple way to consider it at the moment is that the light is completely red-shifted out of the spectrum. One can argue that any rotation will tend to straighten out the space curvature and thereby render extinction unlikely, but, nevertheless, there is a possibility that invisible bodies might exist in the extreme case of a gravitational collapse. Is it possible that the quasars are objects which are near, but have not reached, this stage of condensation?

The same red shift appears to apply to all the spectral lines that are observed from a single quasar. If the red shift were to be gravitational, then it may be argued that the radiation from

different atoms would occur at different levels in the quasar and would therefore be subject to different red shifts. This argument depends upon an extrapolation of the conditions in stellar atmospheres to the very much more extreme conditions at what must be considered the surface of a Schwarzchild-like body. It may be that in these conditions the lines could not be excited at all, for some correspond to transitions which are normally forbidden, and the whole argument is open to some question. The spectroscopists are convinced that the spectra cannot be explained on these lines and the cosmological interpretation of the red shift is currently accepted.

The investigations of radio astronomy have had a profound effect upon cosmology in recent years, for the sources which the radio astronomer detects appear to extend to hitherto unexplored regions of the distant Universe. The Cygnus radio source would still be easily detectable if it were removed to such a distance that the associated nebulosity could no longer be photographed. Apart from the investigation of individual objects such as the quasars, radio astronomy can provide statistical information about the Universe that we live in.

The radio star surveys, such as those performed by the Cambridge interferometer and Sydney Mills Cross instruments, provide data on the intensities of many hundreds of sources. If one plots the number of sources exceeding various intensities against those intensities, using logarithmic axes for the graph, one obtains a very interesting relationship, the log N/log S curve.

In the early Cambridge 2C interferometer survey, an unfortunate error in assessing the effect of confusion of the sources led to some rather wild claims on the interpretation of the steep slope of the log N/log S curve derived from the survey. Much safer criteria are now used and the most recent surveys are probably very reliable. A log N/log S curve for the 178 Mc/s Cambridge synthesis survey is shown in Fig. 20. The shaded portion in the upper part of the curve embraces the limits which may be assigned to the curve when it is extrapolated to weaker sources than those individually recorded, by a statistical technique in which the frequency of occurrence of small deflections is computed as a function of the size of the deflections. The dotted line in the figure has a slope of $-1{\cdot}5$ and this line has a unique significance. It is the slope that would be observed if the radio sources were, on average, equally distributed throughout all space. For sources of the same abso-

lute luminosity this is easy to see, for the intensity decreases as the inverse square of the distance while the number increases with the cube of the distance, for it is proportional to the volume of the sphere of this radius, if one assumes Euclidean geometry.

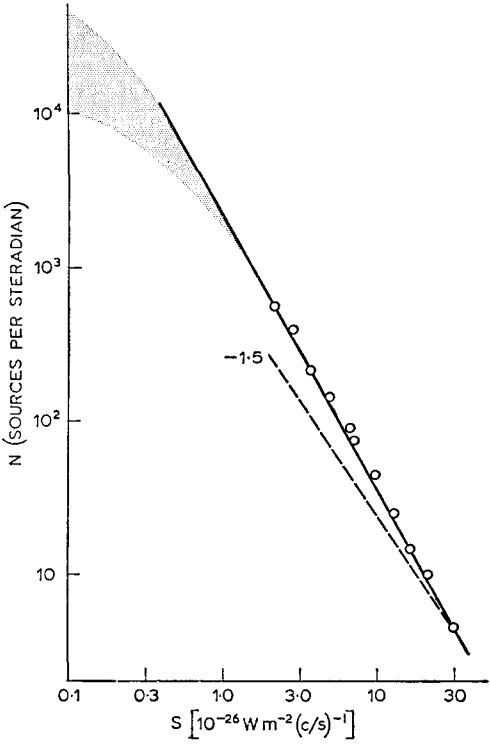


Fig. 20. Log $N/\log S$ curve from the 178 Mc/s Cambridge Survey. The shaded portion is extrapolated by statistical analysis (After M. Ryle)

If the sources are not of the same intensity but are well mixed and free from very large-scale distribution effects, then the argument is more sophisticated but the same result applies. If the slope is less than $-1{\cdot}5$ then there are more sources in the foreground than in the distance, and if the slope is greater

than $-1{\cdot}5$ the reverse applies. Ryle's curve is steeper than $-1{\cdot}5$; if this is a real effect and if weak radio sources are really distant radio sources, it means that the radio sources are more crowded in the distant parts of the Universe. The radio waves from these sources have taken thousands of millions of years to reach us, so that this could also mean that a long time ago the sources were more closely packed than they are now.

Martin Ryle has interpreted this as strong evidence for an evolutionary cosmology and has suggested that it is incompatible with a steady state Universe. Hoyle has pointed out that it is consistent with a steady state Universe in which there are large-scale irregularities. Neither the facts nor the arguments are sufficiently well developed at the time of writing to give a pronouncement on this issue, and we will conclude this chapter with a brief discussion of the elements of cosmology so that the reader may be assisted in interpreting the situation when later works are published.

In cosmology, as in astronomy, the investigator is only an observer. He cannot control or modify the situation which he studies and he must obtain all his information from studying and interpreting every aspect of the radiation, particles, fields and physical laws in his own neighbourhood. From this information the cosmologist endeavours to construct a theoretical model universe, which, if it represented the real Universe, would reproduce the appearance of the cosmos and account for or embrace the major physical laws in our neighbourhood.

The purist might prefer to use only the terms of reference in the previous paragraph but the restriction of the description to that of the single observer hampers the development of the subject. It is the practice of cosmologists to add certain hypothesis which are described as principles. These principles are of such a nature that they cannot *directly* be put to the test, for to do so would mean that we would have either some measure of control over the Universe or else be able to perform measurements in at least two places at the same time, the two places being tens of thousands of millions of light years apart. The principles may be *indirectly* tested by properly using the information available in our own neighbourhood and one is then likely to accept the simplest hypothesis which will account for the observations, for Nature herself bases her designs upon the most elegant simplicity of all.

The prime hypothesis in cosmology is the *cosmological principle*. This principle implies that any two independent and

fundamental observers shall give the same description of the Universe no matter where they be placed within it. Following from this principle the Universe is conceived to be homogeneous and isotropic, it appears to have spherical symmetry to any and every fundamental observer.

The observations of the Universe at large available to a fundamental observer at any instant constitute a *world picture*. A cosmological model in which the successive world pictures of any one fundamental observer are *not* identical is a *periodic* or an evolutionary model. Applying the cosmological principle to a model of this type we assign the same *cosmic epoch* to all fundamental observers who record the same world picture. Cosmological models of this type therefore have a built-in clock and admit the existence of cosmic time. If the model is such that the successive world pictures of any one fundamental observer *are* identical then, as by the cosmological principle all fundamental observers are identical, *all* world pictures seen by all fundamental observers must be identical. In the latter case the model of the Universe, taken in the large, is unchanging, it is a *steady state* Universe and the condition that all the world pictures seen by all fundamental observers should be identical is referred to as the *perfect cosmological principle*.

Of the remaining principles and hypotheses which are of a strictly cosmological nature, reference is frequently made to two. The first of these concerns Olbers' paradox. All around us we see stars and galaxies stretching for a very, very long way. All these objects emit light which, as far as we know, suffers only negligible absorption in its passage over cosmic distances. If we look at the night sky we find that it is dark, yet if the emitting objects continued *ad infinitum* we would expect the sky to be filled with the light from an optically deep distribution of galaxies. It appears that either the Universe itself is not truly infinite or else that it has some property which limits the range of vision of an optical observer and therefore, presumably, the range of total observation of any fundamental observer. Even in the simplest Newtonian cosmology, Olbers' paradox may be overcome by introducing the simple basic postulate of relativity that all observers shall determine the velocity of light to be the same. Each observer then has his own horizon within a 'greater' universe but each sees a similar picture at any one epoch, and, in steady state cosmology, at all times.

The other principle to which reference is often made is Mach's principle. Ernst Mach, famous amongst other things

for the Mach number of supersonic flight, advanced a hypothesis that the seat of all inertial forces lay in the distant stars. This statement is usually slightly re-interpreted at the present day to refer to the distant masses in the Universe. The meaning of this principle may not be so easy to follow, for we tend to accept inertial forces as a basic part of everyday experience. Newton applied his genius to the problem of inertia 300 years ago and formulated the exceedingly profound but simple law that force equals mass times acceleration. He also formulated the equally profound but *separate* law that gravitation exerts a force given by $G\frac{m_1 m_2}{d^2}$ whence an acceleration, g, may be derived. Einstein formally united these laws, at least in part, by postulating the principle of equivalence. One may now argue that this justifies Mach's principle for the inertial force must really be gravitational, but Universal, for all our experiments with gyroscopes and the like appear to show that inertia is independent of the local masses except in so far as actual observations may be affected by the (trivial) effect of the gravitation of nearby objects upon the observing region.

An argument of this sort takes a more restricted view of the principle of equivalence than that principle deserves. It treats it as a principle of identity, that inertia can have no independent existence but must be ultimately attributed to gravitation. If we were to apply the same sort of argument the other way we would have to consider that all gravitational forces have a kinematic origin. The principle of equivalence does not demand identity but only equivalence in so far as no single local observation can decide which of the *alternative* explanations is responsible for the conditions in the local metric.

If one rotates a flywheel then the forces that act on the rim are, according to Mach, the result of the gravitational interaction with distant masses. According to Newton the forces result from kinematic acceleration (though he himself did hint at the possibility of a Mach-like principle), and according to Einstein's principle they may result from either cause. In order to choose between Newton and Mach one has to remove the distant masses in the Universe, a truly formidable problem. The experimentalist cannot remove the masses in the Universe, but the cosmologist can set up a model in which there is no mass and yet by, for example, arranging the annihilation of particles at the right epoch, he can arrange that there shall be rays of light (with no rest mass) to define directions in space. If now he

sets spinning an imaginary rotating system he finds that the metric of that system is just that required to impart forces qualitatively similar to those experienced in a rotating system. The same applies if he kinematically accelerates a small test volume along a straight line, the ray paths of the light which penetrate the volume and which constitute the tools for establishing the local metric, exhibit all the conditions required for the phenomenon of inertia, just as the rays of light around a ponderous mass exhibit the conditions for gravitation. The Universe, on account of its size and of the finite velocity of light, defines a non-rotating system and the conditions for inertia without the introduction of distant masses. W. H. McCrea discussed this condition in his Presidential Address to the Royal Astronomical Society in 1963. There is one point missing in the kinematic argument for inertia and that is whether or not mass itself has any meaning in an empty Universe. At this point the argument could become circular, returning to Mach's principle as the only solution, were it not for the fact the special relativity alone incorporates mass in the famous relation $E = mc^2$. If the fundamental particles are themselves dependent upon the invariance of the velocity of light then their behaviour, if held in the metric of a kinematically accelerated test volume in an empty Universe, would not unreasonably be expected to correspond to that observed. Such a principle is no more profound or obscure than that of Mach. There is good reason to believe that there is indeed a close connection between special relativity, rotation and the mass of fundamental particles. The author discussed the elements of this relationship for the electron in his Presidential Address to the Institution of Electronics in 1964.

The preceding paragraphs have been purposely introduced to enable the reader to see that there is a deep and fundamental problem inherent in Mach's principle, and in the hope that the reader may form his own conclusions as to whether the principle is right, wrong or optional. The principle has been extensively cited in many cosmological works and in particular it is inherent in F. Hoyle's new theory of gravitation, wherein he considers the effects of large scale irregularities in the cosmos.

Most cosmological models give rise to an interesting relationship involving, not the inertial mass, but the gravitational mass, by way of the density of matter and the gravitational

constant. This relationship is

$$G\rho T^2 \gtrsim 1,$$

where G is the gravitational constant,

ρ is the density of matter in the Universe, and

T is the Hubble time—the age of the Universe in an evolutionary cosmology. In most models T is about 10^{10} years and $\rho \sim 10^{-29}$ g.cm^{-3}.

This relationship may be interpreted as showing that the gravitational constant is a function of the age of the Universe. Returning to Newton's laws one could alternatively say that the units of inertial and gravitational mass are not necessarily the same but vary with the passage of time. R. H. Dicke, the inventor of the first switched-reference radio astronomy receiver, has performed an exceedingly accurate experiment using the principle of the Eötvos torsion balance to search for a discrepancy of this type. At the time of writing his accuracy is a few parts in 10^{10} which is not yet quite sufficient to yield a significant result.

The possibility of the units of measurement changing with the passage of time is well recognized in cosmology. One of the more delightful suggestions in cosmology was made by Milne who, in his theory of kinematic relativity, showed that the red shift, the limited observable Universe and many other observable facts followed from the postulate that the unit of time was changing with time. One cannot bottle up time to check this hypothesis for the bottled time ages with the period of storage if mere recordings are taken, so that the test lies in the observation of distant objects, and indeed observation shows that the atoms in galaxies millions of years ago, when the light left them, appear to 'vibrate' at a slower rate and give rise to redder light.

Historically the red shift was predicted before it was observed. It was a prediction of Einstein's general theory of relativity which he applied to a model universe, that any observer would see a universe which was limited by relativity so that information could only propagate over a finite distance and the distant parts would appear to keep a slower time than those nearby, and thus appear red shifted. When Hubble observed that the galaxies did, indeed, exhibit a red shift which depended on their luminosities and therefore their distances, it was hailed as a major success of Einstein's model. Unfortunately the original model was later shown to be in-

correct, but the principles of relativity still form the basis of the more rigorous cosmologies. The extent and the manner in which relativity is introduced varies with individual models. We have already noted that a workable evolutionary model may be constructed simply from the consideration of an expanding cloud of particles and the application of Newton's laws together with the only basic property of relativity, that all observers shall locally measure the velocity of light to have the same value. It is an interesting exercise in this respect to consider the Universe entirely from the viewpoint of classical thermodynamics. If, for example, solid walls could be inserted to make an enclosure at the Hubble radius, would these be subject to a pressure from within, would the central observer's measurements agree with those of the observer at the wall (he would presume that he was at the centre of the Universe and should observe no *local* pressure, though *any* fundamental observer may interpret the apparent general recession of the more *distant* objects as an effect of pressure), and would the temperature everywhere within the enclosure be the same? The matter in the Universe is not smoothed out to the extent that it may really be considered to behave as an ideal gas, for we see the matter concentrated into separate galaxies by local gravitational and magnetic fields. What if one replaces the concept of molecules by the macra-molecules of entire galaxies? These have masses of about 10^{11} solar masses and typical random velocities of the order of 10^8 cm/s so that their kinetic energies are of the order of 10^{60} ergs and the equivalent 'kinetic temperature' of these gigantic systems is then of the order of $1{\cdot}5 \times 10^{76}$°K. This equivalent kinetic temperature may have very little meaning for though the galaxies would collide with the solid boundary wall to impart momentum and give rise to a resultant pressure, the collisions between the galaxies themselves are very unlike those of molecules, for it appears that they may inter-penetrate with little or no effect upon their large scale kinematics and a Maxwell-Boltzman treatment of the macra-molecular gas is questionable. Nevertheless the thought is entertaining, as is almost any concept on a cosmological scale.

CHAPTER EIGHT

TRICKS OF THE TRADE

THE SINGLE AERIAL

A radio telescope is nothing more than an aerial system but the aerial systems used in radio astronomy differ very widely according to the uses that they perform. The variety of tricks that can be performed with combinations of aerials ultimately depends upon the properties of a single aerial.

In our discussion of the single aerial we shall consider the merits and disadvantages of a completely filled aperture; this may or may not imply that the aerial (for example a paraboloid) is equipped with a single element or primary feed and the treatment applies equally to the same aperture filled with a number of elements fed in phase, such as a broadside-collinear array. An ideal paraboloid would be equivalent to a perfectly filled circular aperture if the primary feed aerial at the focus illuminated the whole of the dish so that the electric field over the aperture was constant. In practice this condition is never realized and the illumination invariably falls off towards the rim.

If an aerial is fully filled and the electric field across the aperture is everywhere in phase, then this situation will also appear to prevail if the aerial is viewed from any distant point along the axis in which the aerial is pointing. If, however, the aerial is viewed a little way to one side of the axis the contributions from different parts may not be in phase because the difference in path length to the nearer and further parts may become equal to (or greater than) half a wavelength of the signal frequency. The radiation from parts of the aperture which appear half a wavelength out of phase will cancel and in these directions the resultant signal strength will be less than that along the axis, so that the radiation is beamed or concentrated most strongly in the forward direction. The extent of this concentration and the width of the beam depend upon the size of the aerial and the wavelength of the signal. If the concentration in the forward direction is very great the aerial is said to have a high gain and this aerial (power) gain, g, therefore depends upon the area $\mathbf{A}$, the wavelength λ and the total solid angle of all possible directions in space, 4π. The relation is

$$g = \frac{4\pi\mathbf{A}}{\lambda^2}.$$

g is equal to one for a hypothetical aerial which would radiate (or receive) equally in all directions (an *isotropic* aerial).

The polar diagram of a completely filled aperture is directly analogous to the optical case of the Fraunhofer diffraction pattern of an aperture of similar shape, and may be obtained from the Fourier transform of the distribution of field across the aperture.

$$A(\theta) = \int_{-\infty}^{\infty} f(x)\, e^{2\pi i x\theta}\, dx,$$

where we are considering the projected distribution and polar diagram in the x axis only, $A(\theta)$ representing the amplitude polar diagram or beam pattern of the aerial (the relative amplitude between the signals at various angles, θ, from the axis in the far distance) and $f(x)$ the distribution of the field across its aperture; x is in units of *wavelength* measured across the aperture in the same direction as θ is taken.

The amplitude polar diagram of a rectangular aerial along the x axis is therefore the same as the diffraction pattern of a slit x wavelengths wide and is given by the Fourier transform of a rectangular distribution. This varies as the so-called 'sinc' function

$$A_1(\theta) = \frac{\sin \pi\theta x_0}{\pi\theta x_0}.$$

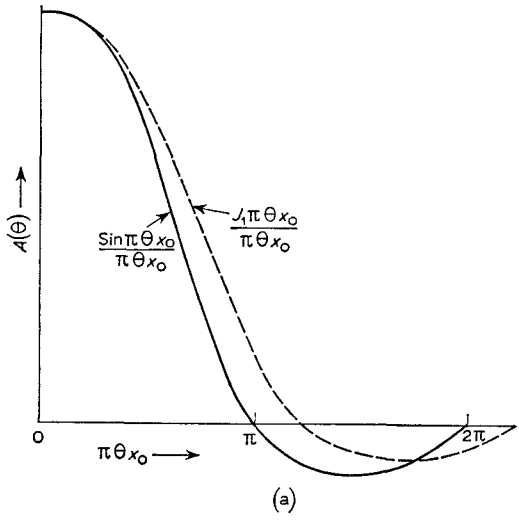
Strictly θ should be $\sin \theta$ but the small angle approximation may be taken if we are considering the case of the very fine beams from large aerial systems.

The amplitude polar diagram of a fully filled circular dish is similar to the diffraction pattern from a circular hole:

$$A_2(\theta) = \frac{J_1 \pi\theta x_0}{\pi\theta x_0},$$

where x_0 is the diameter of the dish in terms of wavelengths, and J_1 denotes a Bessel function. Both $A_1(\theta)$ and $A_2(\theta)$ are quoted in normalized units, that is to say, the gain along the axis perpendicular to the aperture is taken as unity. The transform of the rectangle differs from that of the circular aperture largely in the amplitude of the side lobes and in the position of the minima, the first of which occurs at $\theta = 1/x$ for the rectangle and at $\theta = 1{\cdot}22/x$ for the dish; as x is in wavelengths this may be written $\theta = 1{\cdot}22\lambda/\text{diameter}$.

The transforms we have taken are those relating to the amplitude of the electric vector, whereas the quantity that is



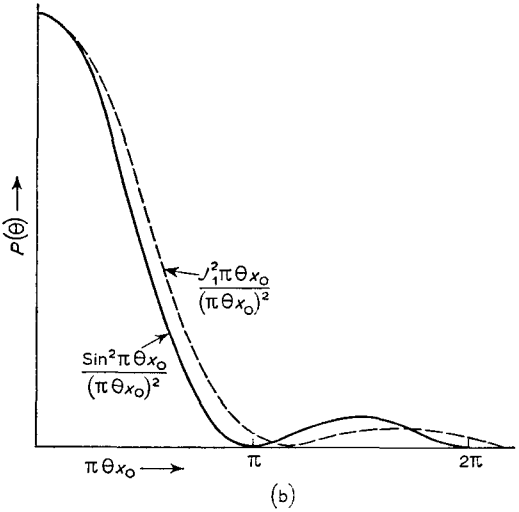


Fig. 21. Directivity patterns of a rectangular telescope aperture (full line) and a circular aperture (dotted)
(a) amplitude (b) intensity
The patterns are symmetrical to the left of the vertical axis

measured is usually the intensity or power output of the receiver. In this case the power polar diagrams, $P_1(\theta)$ and $P_2(\theta)$, are given by the squares of the respective functions. The curves are sharper and, of course, always positive, but the positions of the zeros are not affected. Figure 21(a) depicts the amplitude polar diagrams for a square and a circular aperture, while Fig. 21(b) shows the corresponding intensity plots. The same polar diagrams, or directivity patterns hold for transmission or reception.

The polar diagram of an aerial may be traced directly by a power recorder on the output of its associated receiver if a single point source is moved across the aperture at a great distance from it. The narrower the polar diagram of an aerial, the greater is its ability to resolve detail of small angular structure and for this purpose we shall adopt the usual criterion and consider the resolution of two closely spaced radiating points. Each of these will contribute to the receiver output and the output recorder will trace the sum of their contributions. In general, if the source is incoherent and has an angular distribution of intensity of the form $I(\theta)$, then the output from a receiver connected to an aerial of polar diagram $P(\theta)$ will be given by the sum of the products of the contributions from the different parts of the source with their associated ordinates on the beam pattern as it is swept through the source. This process is known as the convolution of the brightness distribution with the beam pattern, and the output may be expressed as a function $R(\phi)$, of the angular traverse ϕ, of the source through the beam:

$$R(\phi) = \int_{-\infty}^{\infty} P(\phi - \theta)\, I(\theta)\, \mathrm{d}\theta.$$

THE RAYLEIGH CRITERION

It is convenient to define a criterion of resolving power and that introduced by Lord Rayleigh for optical telescopes can be applied directly to radio systems.

The Rayleigh criterion is not always the best criterion but it affords a useful standard of comparison between various systems. It may be stated: *Two point sources may be resolved if the first maximum of the diffraction pattern generated by one source lies on the first minimum of that generated by the other.*

The diffraction pattern here refers to the aerial polar diagram as it is swept through each point, thus two point sources are

resolved when one lies on the peak of the central aerial beam while the second is in the direction of the first minimum. At this separation we have the separate contributions of the two sources to the intensity at the central minimum given by $\frac{\sin^2 \beta}{\beta^2} = \frac{4}{\pi^2} = 0{\cdot}4053$ (for a rectangular aperture). Hence the resultant intensity at the centre relative to the two peaks is approximately 81 per cent (Fig. 22).

If the two sources subtend a slightly smaller angle the trough disappears, whereas the two maxima become entirely separated when the first maximum of one occurs on the second minimum of the other.

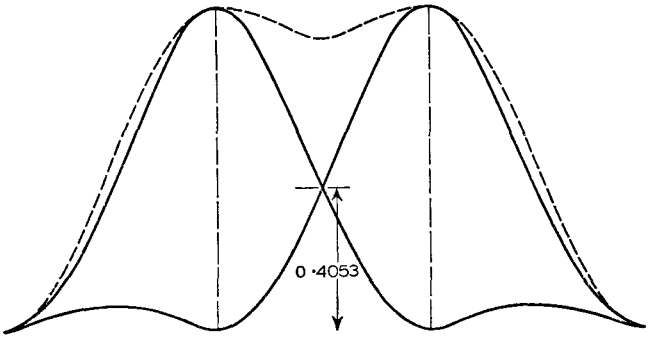


Fig. 22. The Rayleigh criterion of resolution. Individual patterns, full lines; combined pattern, dotted

The Rayleigh criterion yields the result that the sources are resolved when the angle they subtend is $\alpha = \lambda/D$ for a rectangular aperture or $1{\cdot}22\lambda/D$ for a dish. The two maxima are completely separated at $2\lambda/D$ for a rectangle and $2{\cdot}33\lambda/D$ for a dish.

The Rayleigh criterion takes no account of the actual profile of the beam pattern and the criterion becomes meaningless for more unconventional aerials such as a Gaussian distribution, the polar diagram of which is another Gaussian with the first minimum occurring at infinity. Such an aerial is never realizable in practice for it demands that the aerial itself extends to infinity, but in the approximate case, as with other unorthodox distributions, the criterion of resolution must be derived independently for the particular beam pattern. Another difficulty

arises if the sources are of unequal strength and it has been suggested that, in certain circumstances when confusion is severe, up to 20 beam widths should be allowed between sources.

It is apparent from the foregoing that the resolution attainable with a single aerial depends upon its aperture and the wavelength of the radiation that it receives. If the resolution is only required in one dimension, a long narrow aperture may be used to give a fan-shaped beam. It is usually desirable that the resolving power should be high in all directions, and the effective parameter in the design of the system becomes the area of the aperture. A large area also means a large light gathering power so that a system of large overall dimensions is also very sensitive.

If a suitable device such as a harmonic generator is connected in the receiving channel prior to the recorder it is possible to produce an extremely narrow record on the trace. As an example we may select the nth harmonic of the radio or intermediate frequency, amplify this and detect it. The result is to sharpen the recorder response corresponding to the relation,

$$R_n(\theta) = (A(\theta))^n.$$

This sharpening of the recorder characteristic could be erroneously interpreted as an improvement in the aerial polar diagram. In fact the resolving power of the equipment is in no way increased. We may demonstrate this by considering the response to two point sources separated by slightly less than the Rayleigh criterion so that the original response has a broad flat top. By recording the nth harmonic of the signal we do not produce two separate humps but merely sharpen the original function as a whole.

APERTURE AND POLAR DIAGRAM SYNTHESIS

The far-field directivity pattern or polar diagram of a large aerial is dependent upon the dimensions of the aerial system, the way in which the signal is distributed in phase and amplitude across this aperture, and the wavelength of the signal. The precise relationship is expressed by a Fourier transform. Aerials which are engineered in a variety of different ways may therefore have the same polar diagram and will be indistinguishable when viewed from the far distance if they share the same physical conditions for the distribution and the wavelength of the signal across their apertures. The far-field radiation of a paraboloid, for example, could be synthesized by constructing

a collinear broadside array of aerials over a similar circular area and feeding all these aerials in phase by separate cables, but attenuating the feed slightly to the aerials near the circumference to compensate for the weaker illumination of the primary feed of a paraboloid in this region. An artificial satellite passing through the beam pattern of the two aerials could not differentiate between them, and conversely, if the satellite were transmitting, observers on the ground would obtain identical results with both aerial systems. The choice between them would

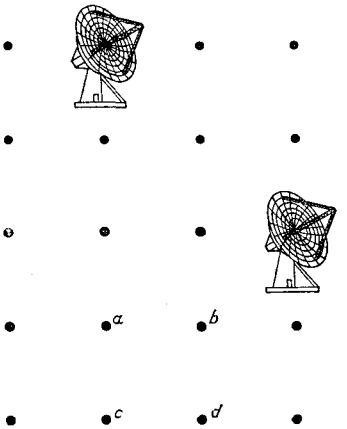


Fig. 23. Aperture synthesis. The combination of readings from different spacings synthesizes the survey that would be obtained with an aerial covering the whole area. Not all possible pairs of readings need be taken—*ab* is the same as *cd* and *ac* the same as *bd*

likely be made upon their engineering merits and production costs alone.

The signals from an artificial satellite change rapidly in direction and also in character with the passage of time, but the signals with which the radio astronomer is usually concerned are fixed in celestial coordinates, and variations in intensity are usually negligible over periods of years. In these circumstances the radio astronomer is very fortunate, for he may synthesize the beam of the paraboloid piecemeal over a long period of time, and, provided that he properly combines all the observations at a later date, he will end up with the same results as would have been obtained with the collinear-broadside array or the paraboloid. It might appear that, in principle, he could go

to the extreme of using only one element of the collinear-broadside array, and, by planting this daily in the position of each of the original elements and recording the amplitude and phase of the signals throughout the day, he could acquire all the necessary information. To record the phase of the signal presents a problem, for the radio sources are incoherent and the phase of the signals is a rapidly and randomly varying function. The solution is to use a second aerial element in conjunction with the first to form an interferometer. The beam pattern of the pair of elementary aerials is a Fourier component of the final synthesized beam while the elementary aerials themselves are obviously components of the aperture. By moving the interferometer pair to every spacing and orientation on the checker board of the original aperture, the identical beam pattern may be built up over a period of time. It will be seen from Fig. 23 that many spacings and orientations are redundant and there is no need to repeat these if the signals from representative spacings are given appropriate weight in the analysis.

The principles of this technique were used in the early measurements of the structure of the cosmic radio sources, but it was later extended to the full by Martin Ryle in the complete synthesis of very large two-dimensional apertures. It is of interest to note that by suitably combining and weighting the component signals it is possible to synthesize aerial systems of an unorthodox nature that would be extremely difficult to engineer. Ryle has extended the technique to that of *super-synthesis* in which the rotation of the Earth is used to carry one elementary aerial system around the other, and by varying the spacing between the aerials a delightful and elegant synthesis may be achieved. Figure 24 shows how a circular aperture may be synthesized. In general the area synthesized is elliptical with axes of D and $D \sin \delta$ where D is the maximum separation of the aerials and δ is the declination of the part of the sky under examination. The information obtained at each spacing is fed into a computer which is programmed to perform the synthesis. It will be noted that, as the beam patterns of the individual elementary aerials may be quite broad, the synthesis can provide information with a synthetic high resolving power over a large part of the sky whilst the aerials remain pointing at a particular celestial coordinate. This information is traded, to some extent, for a high signal to noise ratio on individual sources, but even this may be retrieved by taking a sufficient number of overlapping observations of each part of

the sky. The signal to noise ratio improves as the square root of the number of observations and a simple way to increase it is therefore to observe simultaneously at a number of spacings, albeit at the expense of further elementary aerials and associated equipment. Martin Ryle's instrument at Cambridge incorporates three steerable 60 ft paraboloids one of which is on a very accurate rail track, as the aerial elements. The signal to noise ratio obtained in the final maps produced by this instru-

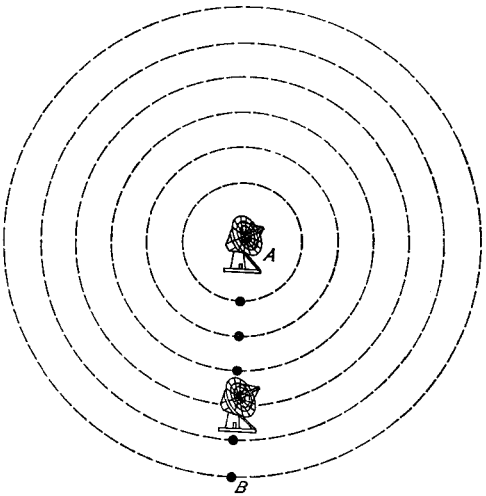


Fig. 24. Supersynthesis is a sophistication of aperture synthesis. The rotation of the Earth is used to sweep out the whole area if the measurements are made along the line AB

ment is equivalent to that which would be obtained from a single aerial having an area of a million square feet.

INTERFEROMETERS

In the previous section it was shown that a large aerial could be reduced into its components, and that the original polar diagram could be reconstructed by summing the contributions from a pair of elementary aerials placed at a sufficient number of points to represent the various spacings and orientations in the original aperture. Although the sources in radio astronomy

are incoherent, the output from the pair of aerials will be periodic, and therefore in a sense it is phase-sensitive, for though the output from a single aerial element as a function of intensity is always positive, that from a pair of aerials oscillates about the mean intensity, following the relative phase angle between the signals in the two aerials as the source passes through the beam. The phase of these *fringes* therefore depends upon the electrical connections and cable lengths, and the relative length of the ray paths from each point in the source to the two elementary aerials. When the signals from both elements are in phase the output is greater than the mean, and when out of phase the output is less. When both elements are similar and the source is a point, the modulation of the pattern is complete and the output intensity varies from zero to twice the power received by a single element. This is analogous to the phenomenon of Young's Slits in optics and is the basis of the simple stellar interferometer. The principle was first applied to an optical telescope by Michelson in 1892 and was adapted to a radio analogue by Ryle in 1947.

If the cables connecting two identical aerials are of equal length the beam pattern follows the pattern of either aerial taken separately, modulated by a system of cosine fringes whose period depends upon the separation of the aerials. In the power record at the output of the receiver the cosine fringes of the signal amplitude become $\cos^2$ fringes in intensity. By measuring the amplitude of these fringes relative to that of the unmodulated output we may derive a function which expresses their visibility or contrast. By determining the visibility at all aerial separations we may compute the distribution of brightness of the source.*

The simple interferometer is the basis of many techniques in Radio Astronomy and it is therefore worthwhile considering its operation from first principles. In Fig. 25, P and Q represent two aerial apertures, having voltage gains A_1 and A_2 respectively, feeding into a common receiver but having feeder losses η_1 and η_2 (voltage attenuation factors) and path lengths differing by a (free space) length of γ. Situated at a great distance above the horizontal plane PQ and at an angle ϕ to the zenith is the radio point source S_1. The signal from the point

* The variation of visibility of the fringes as the two aerials are separated is equivalent to the variation of the modulus of the Fourier transform of the brightness distribution of the source, while the relative shift of the fringes themselves is a measure of the phase component of the transform.

S_1 via aerial P to the receiver, has a voltage amplitude $\frac{A_1\varepsilon_1}{\eta_1}$, where ε_1 depends upon the field strength of the source, and an arbitrary phase ψ. This signal may be written

$$\frac{A_1}{\eta_1}\varepsilon_1 \sin(\omega_1 t + \psi)$$

and the signal from the point S_1 via aerial Q to the receiver,

$$= \frac{A_2}{\eta_2}\varepsilon_1 \sin\left[\omega_1 t + \frac{2\pi}{\lambda}(d \sin \phi_1 + \gamma) + \psi\right].$$

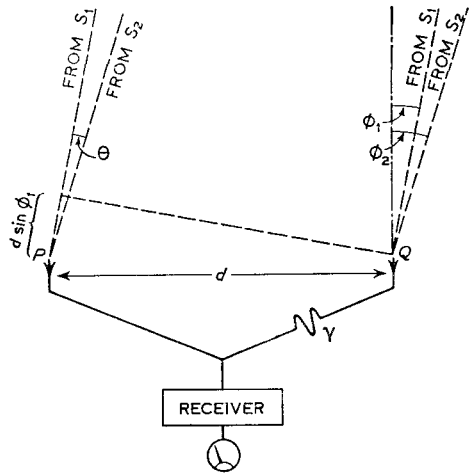


Fig. 25. The Michelson stellar interferometer

The total signal at the input of the receiver due to S_1 is therefore

$$= \varepsilon_1\left\{\frac{A_1}{\eta_1}\sin(\omega_1 t + \psi) + \frac{A_2}{\eta_2}\sin\left[\omega_1 t + \frac{2\pi}{\lambda}(d\sin\phi_1 + \gamma) + \psi\right]\right\}$$

The intensity recorded on a power detector at the output of the receiver

$$= K_1\left|\varepsilon_1^2\right|\left\{\frac{A_1^2}{\eta_1^2}\sin^2(\omega_1 t + \psi) + \frac{A_2^2}{\eta_2^2}\sin^2\left[\omega_1 t + \frac{2\pi}{\lambda}(d\sin\phi_1 + \gamma) + \psi\right]\right.$$

$$+ 2\frac{A_1A_2}{\eta_1\eta_2}\sin(\omega_1 t + \psi)\sin\left[\omega_1 t + \frac{2\pi}{\lambda}(d\sin\phi_1 + \gamma) + \psi\right]\Bigg\}\Bigg|$$

$$= K_1\Bigg|\varepsilon_1{}^2\Bigg\{\tfrac{1}{2}\frac{A_1{}^2}{\eta_1{}^2} + \tfrac{1}{2}\frac{A_2{}^2}{\eta_2{}^2} + \frac{A_1A_2}{\eta_1\eta_2}\left[\cos\frac{2\pi}{\lambda}(d\sin\phi_1 + \gamma)\right]$$

$$+ \text{terms at radio frequency}\Bigg\}\Bigg|$$

In the above, K_1 is a constant dependent upon the detector law and the gain of the receiver.

If the interferometer is symmetrical the expression may be simplified by putting $K = \frac{K_1A_1{}^2}{\eta_2{}^2} = \frac{K_1A_2{}^2}{\eta_2{}^2}$ and the d.c. output of the recorder,

$$I = K\varepsilon_1{}^2\left\{1 + \cos\frac{2\pi}{\lambda}(d\sin\phi_1 + \gamma)\right\}$$

The output of the receiver is therefore fully modulated by fringes, as the path difference between the two aerials varies due to the transit of the source. If the aerial gains and feeder losses are not matched the modulation is not complete and in the case of a distributed source (see below) this may lead to a source of error in the assessment of visibility unless the actual discrepancy is known.

If the source consists of two (incoherent) points the output recorder has the further contribution from point S_2,

$$I_2 = K\varepsilon_1{}^2\left\{1 + \cos\frac{2\pi}{\lambda}(d\sin\phi_2 + \gamma)\right\}$$

If the average signal strengths of the two points are the same $\overline{\varepsilon_1{}^2} = \overline{\varepsilon_2{}^2} = \overline{\varepsilon^2}$ and we may write for the recorder output,

$$I_T = 2K\overline{\varepsilon^2}\Bigg\{1 + \cos\frac{2\pi}{\lambda}\left[\left(\frac{\sin\phi_1 + \sin\phi_2}{2}\right)d + \gamma\right]$$

$$\cos\frac{\pi d}{\lambda}(\sin\phi_1 - \sin\phi_2)\Bigg\}$$

If s_1 and s_2 are close together, separated by the small angle θ near the direction ϕ, I_T is given by

$$2K\overline{\varepsilon^2}\left\{1 + \cos\frac{2\pi}{\lambda}(d\sin\phi + \gamma)\cos\left(\frac{\pi d}{\lambda}\theta\cos\phi\right)\right\}$$

Where we have adopted the small angle approximations

$\sin\theta = \theta$ and $\cos\theta = 1$. In both expressions the first cosine component shows the production of the fringe pattern as the star and aerial systems are relatively moved. The second cosine is a modulating factor depending upon the angle subtended between the two points at the aerial system. The fringes disappear when this term reduces to zero, that is when

$$\frac{\pi\theta(d\cos\phi)}{\lambda} = \frac{\pi}{2}$$

$$\text{i.e. } \theta = \frac{\lambda}{2d\cos\phi}$$

where $d\cos\phi$ is the projected separation of the aerials P and Q as seen from the direction ϕ and θ is the angular separation of the points S_1 and S_2 as seen from the aerials.

This is the resolution criterion for a simple interferometer.

If the source is uniformly extended between the limits $-\Theta/2$ and $\Theta/2$ the resultant recorded intensity

$$I_T = K\overline{\varepsilon^2}\int_{-\Theta/2}^{\Theta/2}\left[1 + \cos\frac{2\pi}{\lambda}(d\sin(\phi+\theta) + \gamma)\right]d\theta$$

$$= K\overline{\varepsilon^2}\left(\Theta + \frac{\lambda}{\pi d\cos\phi}\cos\frac{2\pi}{\lambda}(\gamma + d\sin\phi)\sin\frac{\pi}{\lambda}\Theta d\cos\phi\right)$$

$$= K\overline{\varepsilon^2}a\left(1 + \frac{\sin(\pi\Theta d\cos\phi/\lambda)}{(\pi\Theta d\cos\phi/\lambda)}.\cos\frac{2\pi(\gamma + d\sin\phi)}{\lambda}\right)$$

Where $\overline{\varepsilon^2}d\theta$ is now the mean square voltage from an infinitessimal element of the source.

By comparing the intensity at fringe maxima and minima we may introduce Michelson's visibility criterion

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

$$= \frac{2K\overline{\varepsilon^2}\Theta\sin\pi\Theta d/\lambda}{\pi\Theta d/\lambda}.\frac{1}{2K\overline{\varepsilon^2}\Theta} = \frac{\sin\pi\Theta d/\lambda}{\pi\Theta d/\lambda}$$

Where we have taken the star in transit at $\phi = 0$.

(This result represents the modulus of the Fourier transform of a linearly extended source such as a rectangular aperture, without reference to the phase of the transform. The reason for the absence of information about the phase is that in this simple analysis we have taken a symmetrical source and assumed that its centre of radiation is on the centre line of the system.)

Interferometers of the type we have just investigated are the radio frequency equivalent of the optical interferometer (Fig. 4(a)) proposed by Michelson and used for the measurement of

the angular diameter of the moons of Jupiter and such stars as Betelgeuse and Arcturus. It will be noticed that the first disappearance of the fringes from two points of angular separation θ occurs when $\theta = \lambda/2d$. Thus the resolution provided by two slits situated at either edge of a rectangular aperture is twice as great as that of the aperture itself.

The Michelson-type interferometer was successfully used by Ryle for the early surveys at Cambridge. The technique has now been somewhat supplanted by the introduction of the switched interferometer (see page 117), but the original method retains the advantage of direct measurements of visibility provided that the system is designed and used with care.

An interferometer system originally used by Pawsey and Bolton for measurements in Australia is based on the principle

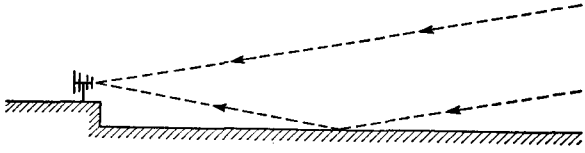


Fig. 26. The sea interferometer

of Lloyd's mirror, in which the rays from the source fall upon the receiving element by two paths, one direct and the other by reflection from a large plain reflecting surface. The aerial system was arranged at the top of a cliff and the sea formed the plain reflecting sheet (Fig. 26). In practice the aerial system may be designed to permit the use of higher gain and narrower beams, and in this form it is also amenable to conversion to a switching technique. The analysis of the sea interferometer is basically similar to that of the Michelson, apart from the phase change introduced by the reflection.

The simple interferometer is a powerful method for increasing the resolving power of an aerial system but it has its limitations. In order that we may appreciate the necessity for more complex methods we shall enumerate some of the disadvantages which limit the use of the simple system.

DISADVANTAGES OF THE SIMPLE INTERFEROMETER SYSTEM

(1) *Response to background intensity.* The Michelson stellar interferometer in its simple form responds to the total power

of large resolved features of the sky. These do not give rise to fringes but may cause very large changes in the mean level of the signal. If one wishes to amplify the output in order to obtain large fringes from a small and weak source, the excursions of the recording system resulting from the background radiation may prevent the retention of the fringes on the recording scale.

(2) *Background confusion.* In order to measure the variation of visibility of the fringes it is necessary to obtain an accurate measure of the normal visibility at very short spacings or of the power from the source. This may be extremely difficult when observing sources of small intensity, as a result of the uncertainty in the contribution of the background radiation. This effect is pronounced where faint sources are situated in a region of strong galactic radiation, and is again encountered when we wish to investigate the details of a source whose broad structure is resolved. With very long baseline interferometers it is often impossible to obtain running calibrations at short spacings.

(3) *The matching of the system.* If an estimate of the visibility of the fringes is to be made it is very desirable that the gains of both of the aerials and the losses in the two feeder runs shall be identical. If this is not so the modulation of the total power from the source is never complete and an erroneous measure of the visibility may be obtained. It is also essential that the electrical lengths of the aerial feeders shall be identical or the estimation of the position of the source will not be correct. If the Michelson criterion of visibility is not used it is, of course, possible to calculate the brightness distribution from a true knowledge of the actual aerial and feeder distribution, but this is subject to some error and is not, in general, to be recommended.

(4) *The attenuation of the feeder system.* If the two aerials are separated by more than a few hundred wavelengths the attenuation of the radio frequency signals in the cables or feeders connected to each aerial becomes so large that the signal to noise ratio of the equipment appreciably suffers. The effect depends upon the nature of the feeder system and on the frequency of operation, but is a serious limiting factor to the available resolution of the cosmic sources. If an attempt is made to overcome the feeder loss by the use of separate pre-amplifiers these must be accurately matched in gain and phase, for otherwise the fringe modulation will not be a true representation of the visibility or position of the fringes.

(5) *Azimuth restriction.* It is not usually practicable to tilt the baseline of an interferometer in the vertical plane, and the transit of the source is employed to produce the fringed pattern on the output recorder. If the baseline is due East and West this produces a fully fringed pattern, but, as the baseline is rotated in azimuth, the period of the fringes will be increased until, for a baseline orientated due North and South, the fringe frequency will be zero at transit.

In these circumstances the source may occupy any position on the lobes of the polar diagram and it is not possible to obtain a direct reading of the visibility at any point along the baseline. It is possible in principle to arrange a continuously variable length of baseline to produce fringes and the source distribution in the direction due North–South may be computed from the variation of these fringes with the aperture spacing. It is also possible, and more simple, to rotate the phase between two fixed North–South aerials, thereby effectively causing the lobes to 'rotate' through the source while retaining the same baseline and resolving power. This trick is not inherent in the principle of Michelson's stellar interferometer and it will be discussed later in this chapter.

(6) *Stability of amplitude.* The requirements for amplitude stability are twofold. It is essential that the amplitude stability of the aerials and feeder system be exceptionally high as required under (3). It is essential that the stability of the remainder of the equipment be such that the sensitivity does not change appreciably during a single observation. It is also of considerable practical advantage if the long term stability is such as to maintain the output continually in proportion to the total input. It is not possible to arrange a continuous automatic gain control on a simple interferometer as this would iron out the deflections we wish to observe; it is, however, possible to use a switched reference or servo control system if it is desired. A.G.C. may be used on a phase-switched interferometer, as explained in a later section.

(7) *Stability of phase.* If we desire to measure either the true source distribution or the accurate position of the source it is essential that the phase stability of the equipment shall be very high. Even with high quality feeder systems the phase stability is subject to atmospheric changes, and local heating of part of one of the feeder limbs may cause appreciable changes in the phase of the signal, and hence the apparent position or structure of the sources. The use of pressurized and buried

cables is to be recommended where the installation and finances permit. Simple interferometers on extended baselines are often rendered phase-insensitive by the phase instability in the long signal paths and the accumulated phase errors in their electronic circuits and local oscillator paths.

(8) *Stability of Frequency.* The effect of a small change in the frequency of the recording receiver is equivalent to a change in the separation of the two apertures and causes a corresponding motion of the fringes. The effect is more pronounced at large spacings and may be calculated in a similar manner to that used for bandwidth dispersion (see 10 below). If the feeder system is symmetrical the requirement is only that the stability shall be a small part of the bandwidth, but if the two limbs of the interferometer are not balanced the restrictions

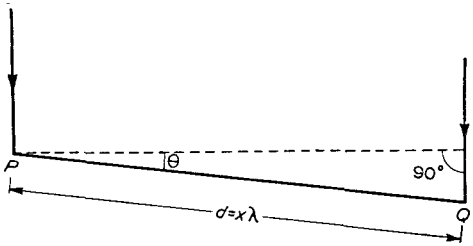


Fig. 27. Ray paths applicable to dispersion

are extremely severe; a frequency variation of 0·01 per cent in a feeder or link system of only 10,000λ represents a phase shift of 30°.

(9) *Inhomogeneous ionosphere.* The occurrence of scintillations in the ionosphere can cause serious error in the amplitude and phase of the signals received at the two apertures, particularly when the baseline is extended beyond the distance over which the fluctuations are correlated. Fortunately severe scintillations are not a regular occurrence: they do not seriously affect daytime observations and their effect is less marked at higher frequencies.

(10) *Bandwidth Limitations.* The bandwidth limitation is inherent in any extended aerial system irrespective of its design. The receiver pass-band defines the effective source spectrum when the latter is assumed inherently 'white'. Thus the time over which the radiation may be considered coherent is determined by the decrement of the circuitry and the actual

time coherency may be calculated from the Fourier transform of the pass-band of the receiver.

The effect of a finite pass-band is to produce a frequency dispersion which is reflected into the effective aperture of the aerial system. In Fig. 27, the path difference between the two rays arriving at points P and Q of an aerial aperture is $d \sin \theta$ and the delay in time is $t = (d/c) \sin \theta$ where c is the velocity of light. Substituting $d = x\lambda$ and $c/\lambda = f$ we have

$$t = \frac{x}{f} \sin \theta \quad \text{or} \quad x = \frac{tf}{\sin \theta}.$$

Hence the frequency dispersion $\delta f/f_0$ gives rise to a corresponding equivalent dispersion in the baseline of $\delta x/x$.

Now δf has all values within the receiver bandwidth, from which it is apparent that the resulting dispersion in x is tantamount to a re-definition of the aerial aperture by a function depending upon the total extent of the aerial system and the receiver bandwidth. The resulting effect upon the aerial beam is to multiply the aerial beam pattern by the Fourier transform of the bandwidth of the equipment. Hence for an achromatic system

$$R(\phi) = \int I(\phi - \theta)\, P(\theta) \int B(f) e^{2\pi i \theta f} df d\theta,$$

where $B(f)$ is the power spectrum of the receiver relative to zero frequency and the remaining symbols are as defined on p. 103. If we denote the transform of the receiver frequency response by $B(\theta)$ we have

$$R(\phi) = \int I(\phi - \theta)\, P(\theta)\, B(\theta)\, d\theta,$$

from which it is apparent that if the baseline is extended the polar diagram is limited by the transform of the receiver frequency response. This is equivalent to the narrow range of white light fringes obtainable with an optical interferometer.

THE PHASE-SWITCHED INTERFEROMETER

The switching technique introduced by Ryle was a considerable advancement in interferometer design. In its original form it immediately enabled the sensitivity and stability to be increased very considerably. The principle involved may also be used to supplement the operation of the more complicated systems working on very long baselines and to perform a variety of tricks with aerial apertures as will be explained later.

An elementary block diagram of the phase-switched interferometer is shown in Fig. 6, Chapter Two. The main feature is the switch which alternately connects the two aerials in phase and antiphase at a low frequency (~ 30 c/s) determined by the switch driver unit. The signals are passed through a receiver, which may be equipped with an automatic gain control, into a detector from which the 30 c/s components are amplified by a selective amplifier and demodulated in a phase-sensitive rectifier or synchronous detector which is equivalent to a synchronous reversing switch and is controlled by the original switch unit. The output of this demodulator feeds the recording meter.

When the two aerials are connected in phase the power polar diagram of the combined aerial system will be of the form,

$$P_1(\theta) = A_1^2(\theta) + A_2^2(\theta) + 2A_1(\theta)A_2(\theta) \cos x\theta,$$

where $A_1(\theta)$ and $A_2(\theta)$ are the amplitude polar diagrams of the two component aerials.

When the aerials are in antiphase the combined aerial polar diagram will be,

$$P_2(\theta) = A_1^2(\theta) + A_2^2(\theta) - 2A_1(\theta)A_2(\theta) \cos x\theta.$$

If we now apply the general formula

$$R(\phi) = \int I(\phi - \theta)\, P(\theta)\, d\theta,$$

we have for the output at the square law detector from the source function $I(\phi - \theta)$,
(i) for the aerials in phase:

$$\begin{aligned} R_1(\phi) &= \int I(\phi - \theta)\, P_1(\theta)\, d\theta, \\ &= \int I(\phi - \theta) A_1^2(\theta) d\theta + \int I(\phi - \theta) A_2^2(\theta) d\theta \\ &\qquad + 2\int I(\phi - \theta) A_1(\theta) A_2(\theta) \cos x\theta d\theta, \end{aligned}$$

(ii) for the aerials in antiphase:

$$\begin{aligned} R_2(\phi) &= \int I(\phi - \theta) P_2(\theta) d\theta \\ &= \int I(\phi - \theta) A_1^2(\theta) d\theta + \int I(\phi - \theta) A_2^2(\theta) d\theta \\ &\qquad - 2\int I(\phi - \theta) A_1(\theta) A_2(\theta) \cos x\theta d\theta. \end{aligned}$$

It will be observed that the first two integrals in each expression refer to the signal power received from the independent aerials and that they remain of the same sign and form in both the expression for $R_1(\phi)$ and $R_2(\phi)$. The last integral in each expression represents the behaviour of the fringe pattern between the two aerials when illuminated by the distributed source $I(\phi - \theta)$ and this is of opposite sign in the two expressions. Thus the final integral represents an alternating component at the switch frequency modulated by the interferometer fringes. This alternating component is able to pass through the tuned 30 c/s amplifier to the phase-sensitive rectifier, while the first two integrals which would form the 'carrier' deflection of the source in a simple interferometer, are rejected in these same circuits.

The phase-sensitive rectifier therefore gives an output of the form,

$$R_0(\phi) = K\int I(\phi - \theta)A_1(\theta)A_2(\theta)\cos x\theta d\theta,$$

where the term $A_1(\theta)A_2(\theta)$ defines the envelope of the output and the term $\cos x\theta$ determines the fringes within the envelope.

If the aerial spacing is large compared to the aerials themselves we may treat $A_1(\theta)A_2(\theta)$ as constant and hence,

$$R_0(\phi) = KA_1(\theta)A_2(\theta)\int I(\phi - \theta)\cos x\theta d\theta.$$

The first equation for $R_0(\phi)$ is the more general expression and should always be used if the analysis covers more than a very small part of the whole pattern.

The analysis shows that the output of a switched interferometer is of the form shown in Fig. 6(b), Chapter Two, consisting of a set of fringes enclosed within the envelope determined by the product of the separate aerial amplitude beam patterns, that is to say that the resultant equivalent aerial gain is the geometric mean of the power gains of the two component aerials taken separately. An aerial with a power gain of 1000 connected to interfere with a small Yagi of gain 10 would yield a combination in which the gain would be 100 but the resolving power would be high and dependent upon the spacing between the aerials. The period of the fringes is dependent upon the aerial spacing, x wavelengths, while the nett amplitude and position of the whole pattern are determined by the source intensity function $I(\phi - \theta)$. The 'carrier' pattern formed by the

total power of the source and the background radiation does not appear in the output, so that the Michelson visibility criterion may not be applied directly.

In order to compute the brightness distribution of a source it is necessary that the whole equipment be maintained constant while the value of $R_0(\phi)$ is established at all spacings from zero outwards. Fortunately the method enables the receiving equipment to be heavily subjected to automatic gain control with a period which is long compared to that of the switch frequency, so that the stabilization of the receiver is high. This is not possible with the simple interferometer, for the a.g.c. would remove the pattern itself. It will be apparent that this method may be used even if the two aerials have vastly different apertures. The aerials may be equipped with pre-amplifiers which may be used to boost the signal power at either aperture and compensate for the loss incurred in the feeder system at longer baselines.

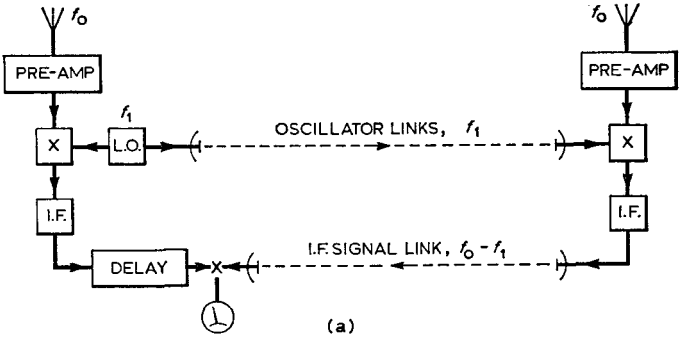
The main source of error in the computation of source distributions, where the limit is not set by confusion or the noise level of the trace, is in the variation of the gain of the movable aerial and its associated pre-amplifier when removed to different sites. This difficulty was overcome by F. G. Smith for the special case when two aerials, the gains of which are invariant, are used at either extreme of the baseline. A more general method based on a system of relative measurement was devised by the author and requires no assumption of constancy in the equipment prior to the ultimate mixing of the channels. This method will be described in a later section, but the method used by Smith will now be briefly discussed.

The two aerials A_1 and A_2 and their associated pre-amplifiers are fixed at either extreme of the baseline. The movable aerial A_3 and its pre-amplifier is used in conjunction with either aerial in turn for a period of a few minutes, or sufficient time to provide a record of a few fringes. The ratio of the amplitudes of the two records is first obtained when the aerial A is exactly halfway between the two large aerials and this serves to calibrate the equipment for the relative gains of the two channels A_1 and A_2. Thereafter the aerial A_3 may be moved to any position between A_1 and A_2, and after the ratio of the new amplitudes has been corrected by the constant derived at the midway position, the resulting ratios give the relative amplitude of the Fourier transform of the source distribution at the two spacings. It will be observed that the characteristics of the movable aerial

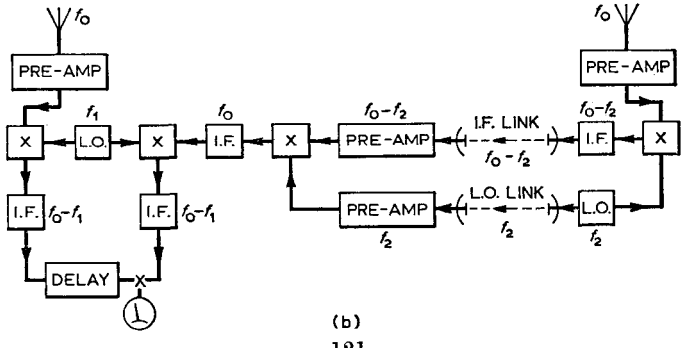
and its pre-amplifier cancel out in the method of measurements and may permissibly vary considerably at different sites.

RADIO LINKED INTERFEROMETERS

When the separation of the interferometer aerials is very large, separate superheterodyne receivers may be employed at each aerial, and the intermediate frequency signals are then united in a common switched detector and recorder system. In order to preserve the coherency of the system it is necessary to feed the same local oscillator into both channels. The local oscillator and intermediate frequency signals may be relayed by cables or by radio frequency links, though in the latter case it is again essential that, if any additional frequency changing or modulating systems are used, the resulting incoherency is removed by homodyning the same oscillator with the signal at the receiving end. In all cases it is necessary to balance the



(a)



(b)

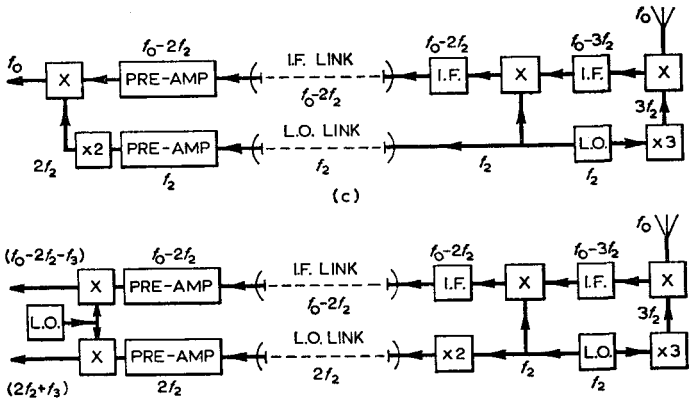


Fig. 28. Coherent radio links. The more complex arrangements in (c) and (d) show only remote station and link circuits. These circuits are less troubled by feedback instability and (d) shows how the L.O. link may be boosted prior to mixing by using a further oscillator. In this case the original frequency, f_o, may be obtained by amplifying and mixing the two outputs. It is usually preferable not to reconstitute f_o but to inject an oscillator into both limbs in order to amplify coherently at a lower frequency

delay in the relay channel by inserting a similar delay in the direct channel.

Figure 28(a) shows the simplest form of a coherent radio linked interferometer. The local oscillator is shown at the home station, but it may equally well be at the distant site. An alternative method using a homodyne link is shown in Fig. 28(b). As this method places no restrictions on the link frequencies, it is the preferable method for use where more than one distant station is employed and it is desired to keep the link channels separate. It is apparent that many alternative arrangements are available according to the choice of frequencies and the selection of either upper or lower side bands.

Both the methods shown in Fig. 28(a) and (b) suffer from the practical difficulty of achieving a high gain in the i.f. strip at the distant station, while preventing feedback from the transmitting aerial into the early stages of the i.f. strip. The method shown in Fig. 28(c) obviates this difficulty by referring the achievement of gain to an intermediate strip well removed from the transmission and reception frequencies. Figure 28(d) is a

more sophisticated version in which the received link signals are boosted by an additional oscillator. The output signals must be combined after further amplification to remove the effect of the additional oscillator. It is not necessary to reconstitute the frequency f_0 provided that the final frequency bands in both limbs of the interferometer are maintained coherent by arranging that both limbs ultimately suffer the same frequency conversions. A possible method of obtaining a good noise factor while also forming a simple radio link is to use the principle of up-conversion on a variable reactance crystal. If the 'pump' frequency f_1 is very much higher than the signal frequency f_0 a very good noise factor and gain are concurrently achieved.

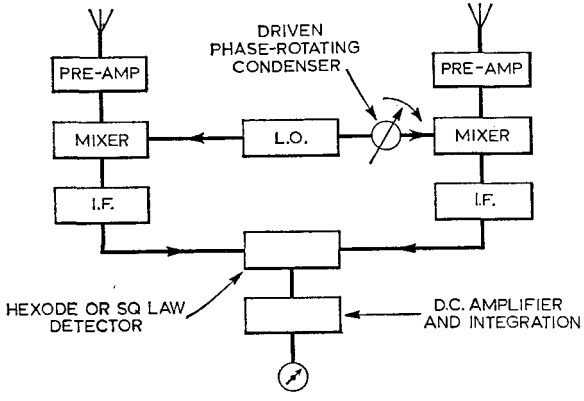
In all radio linked interferometers it is essential that the signal channel shall be linear throughout, and this invariably forbids the use of class C amplification for the final stages of the signal transmitter though it is in order for the reference oscillator channels.

It should be stressed that all high resolution interferometers of the complex electronic patterns described above are very questionable instruments when used for the measurement of detailed structure, unless used with some system for the measurement of total power in each channel or with a third aerial in either the method used by Smith or the author. The measurement of the amplitude of the transform must always assume complete constancy of the equipment and the relay paths. Even with the application of heavy a.g.c. this requirement is too stringent for the measurement of accurate readings. The second and more serious disadvantage is that the information relating to the phase of the transform is completely lost unless it is possible by some means to ascertain the exact phase rotation of the local oscillator and relay paths at various aerial spacings. The problem of ascertaining the phase was originally solved by the author and his technique will be described in the following section.

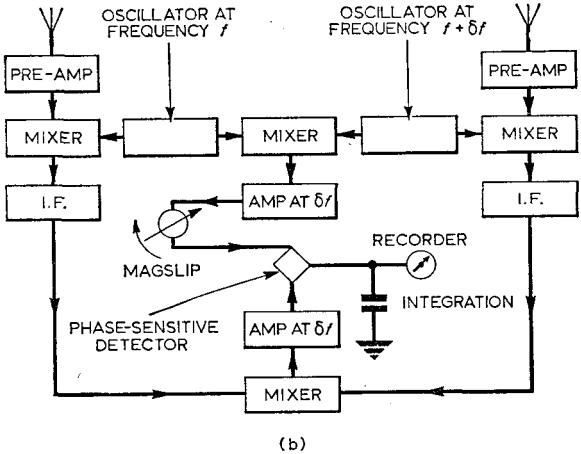
A modification of the switched interferometer is due to Hanbury Brown.* The switching method relies upon the transit of the source to produce the fringes, but if the baseline of the interferometer is due North and South no fringes are formed and the output will correspond to the difference in intensity between two indeterminate points spaced 180° on the fringe

* A similar but less elegant method in which the phase rotation is discontinuous was originally used by Little & Payne-Scott in Australia.

system. Hanbury Brown's method is to rotate continuously the fringe system by smoothly and continuously rotating the phase of the local oscillator in one channel of the equipment; the basic



(a)



(b)

Fig. 29. Rotating lobe interferometer
(a) basic technique using tuned quadrant capacitator in local oscillator lead (Jennison and Latham, 1954)
(b) Hanbury Brown's technique using independent local oscillators

method illustrated in Fig. 29(a) was used by Latham and the author. The more elegant form in Fig. 29(b) was originally used by Hanbury Brown, Palmer and Thompson. It uses independent local oscillators, the beat frequency between which provides the reference signal for the phase-sensitive detector.

PHASE-SENSITIVE INTERFEROMETRY

If the path lengths of an interferometer are equalized and if the equipment is phase and frequency stable the Right Ascension of a point source or of a partially resolved symmetrical source may be obtained by timing the transit of the source through the central lobe of the aerial beam. The declination of the source affects the rate at which it sweeps through the lobes. If a source is at the pole it appears to remain stationary whereas at other declinations,

$$T = t_0 \sec \delta,$$

where δ is the declination and t_0 is the periodicity of the fringe system for a source in the equatorial plane.

With the measurement of the position of a partially resolved asymmetric source, an interferometer measurement taken at one separation of the aerials may not agree with that taken at another, for a large distributed component of the source may first be resolved at short baselines and the apparent position may shift towards a small bright feature as the separation is increased.

The different positions of sources or of parts of a source in their behaviour upon the lobes of the polar diagram of an interferometer are equivalent to distributions of incoherent signals of the appropriate strengths at various path differences to the two aerials, and this in turn is equivalent to a distribution in phase for the various component cosine fringes which add together in the output of the interferometer to give the resultant fringe system from the source at each aerial spacing. In order to determine the true distribution of brightness it is therefore essential that the phase of these signals shall be accurately retained.

There are two ways in which the phase information may be retained. One of these is *absolute* and it then gives the correct celestial position of the source or sources together with their structure up to the limits of the phase-stable resolving power. This technique is simply the application of extremely high standards to the siting of the aerials and to their interconnection by cables, which either do not vary in electrical length

or which can be continuously and accurately measured. Practical difficulties usually limit this technique to resolving powers greater than about one minute of arc.

The considerable extension of an interferometer baseline which is required to resolve detail subtending an angle of less than, say, 1″, would incur the loss of phase information in a two element interferometer. Where the task of the interferometer is purely to establish an order of magnitude for the effective diameter of a source this loss is not a serious disadvantage; but if it is desired to determine the structure of a source or the relative distribution of a group of sources, a knowledge of the true phase of the signals is important. If we do not measure the phase we implicitly assume that the source distribution is symmetrical. Our present knowledge of the sources indicates that such an assumption is often unjustified. There is fortunately a solution to the problem if we are prepared to accept a relative phase measurement restricted to the contributions from the bright objects within the field of view. This means that we obtain a very good unambiguous picture of the small part of the sky that we are examining but we cannot tell precisely where that part is, that is to say we sacrifice absolute position.

The measurement of the phase component of a brightness distribution by timing the transit of the fringe system is of no avail when the interferometer is of the complex nature dictated by the need to achieve a very high resolving power. The combination of phase errors in the Earth's atmosphere, in the electronic equipment, in the local oscillator and intermediate frequency signal link paths, together with unequal signal paths when the terrain is not level, renders the phase of the signals indeterminate at aerial spacings in excess of a few thousand wavelengths. In order to retrieve the phase sensitivity and thus to possess all the information required to reconstitute the original source structure, the author devised the three aerial technique in 1953, and a year or two later used it to perform the first unambiguous measurements of the structure of the Cassiopeia and Cygnus sources. It has been applied by Twiss, Carter and Little to the measurement of a number of sources in the southern hemisphere. The method is simple but powerful, for many errors can be removed. The principle is outlined below.

Three aerials A, B, C (Fig. 30) are aligned along the axis in which the measurements are to be conducted. Two of the aerials are equipped with superheterodyne receiver-transmitters which enable the received signals to be converted to inter-

mediate frequencies which are transmitted with the associated local oscillator frequencies to the third station where they may be reconstituted to form the original signals. Suitable delays are placed in two of the channels to compensate for the path

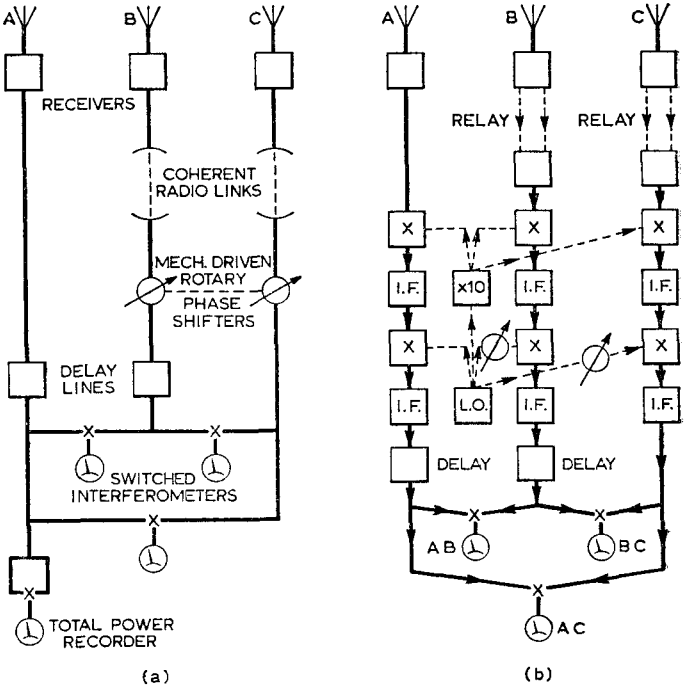


Fig. 30

(a) The principal features, and

(b) basic block diagrams of the phase-sensitive radio linked interferometer. The upper blocks of channels B and C are similar to the radio links in Fig. 28(d)

The phase rotators are driven by separate variable gears from the same motor shaft

differences of the relay links. The apparatus at the third station now has available three signals—they are still entirely separated in the equipment—each of which is a replica of the signal received from the source at the three sites, but each has suffered an unknown phase rotation.

The signals are now brought together into three separate interferometer channels so that the fringes produced between A and B, B and C and A and C may be displayed independently. A comparison of the phase of the fringes produced in the third interferometer (A and C) with those in the first two, directly yields the relative phase change due to the source as the aerials are moved apart to greater distances. Unknown path differences and phase errors resulting from atmospheric refraction or vagaries of the equipment and radio links cancel out in the comparison of the readings.

If the observations are commenced at small spacings, the central station, B, may conveniently be transplanted to twice the spacing A–C and the observations repeated with the station C now in the centre. In this manner the phase of the transform may be mapped up to any aerial separation where the visibility of the fringes is sufficient to measure their phase in the presence of noise.

If the scanning of the fringes is performed entirely by the rotation of the Earth it is preferable for the central aerial to be situated exactly midway between the other two. In practice this requirement is difficult to satisfy because of the nature of the surrounding countryside. If the central aerial is not midway between the outer pair the frequency of the fringes between the pairs will differ. A *direct* phase comparison is rendered impossible but it may be shown that the relative phase contribution of the source may be obtained by subtracting from the argument of the A–C fringes the sum of the arguments of the A–B and B–C fringes. In the author's interferometer the scanning of the fringes was performed by including continuously rotating phase shifters in series with the preset phasing controls from the second local oscillator (Fig. 30(b)). The angular velocity of the rotating phase shifters was capable of fine adjustment to any selected speed and it was possible to scan the fringes from aerials A and B at a similar rate to those from B and C even where the spacings A–B and B–C were vastly different.

The use of three signals in this manner has another property; by adjusting the gain control in one channel so that the amplitudes of the fringes between the central aerial and each of the outer pair are equalized, we also equalize the signal gains of the two outer channels. This property may be used to considerable advantage in the phasing of very extended multiple element systems. A method is also available for determining the

relative amplitude of the transform if the gain of only one channel is known, as distinct from the need for constant gain in two channels in the method used by Smith.

If station B is moved to various positions between A and C the ratio of the fringe amplitudes A–B and B–C may be employed to yield the amplitude of the transform. It is not necessary to commence with station B midway between A and C provided that the values for the transform amplitude at spacings up to A–B or B–C, whichever is the greater, have been determined from a series of readings taken with C at a closer spacing. The values previously obtained may be used to scale the fringe amplitudes as B is moved in the new series of readings. The experimental procedure therefore follows a very similar pattern to that used for the determination of the phase of the transform, so that both measurements may be performed concurrently.

NON-LINEAR INTERFEROMETERS

In the preceding sections we have considered the direct interference of the radio frequency signals *prior* to a single common non-linear device, the signal detector, which is provided for the measurement of signal intensity.

The first interferometer in which non-linear elements were introduced in the two channels before they were mutually interfered was the intensity interferometer, suggested by Hanbury Brown and constructed by Das Gupta and the author at Jodrell Bank in 1950. It was the first interferometer to resolve a small radio source into more than one component, and its optical analogue is now a powerful tool in optical astronomy for the measurement of the small angular diameters of O and B stars. It is capable of extremely high resolving power and is not affected by scintillations. It has the inherent disadvantage of poor signal to noise ratio when the power from the detail under investigation is less than the power of the background noise and it is not phase-sensitive.

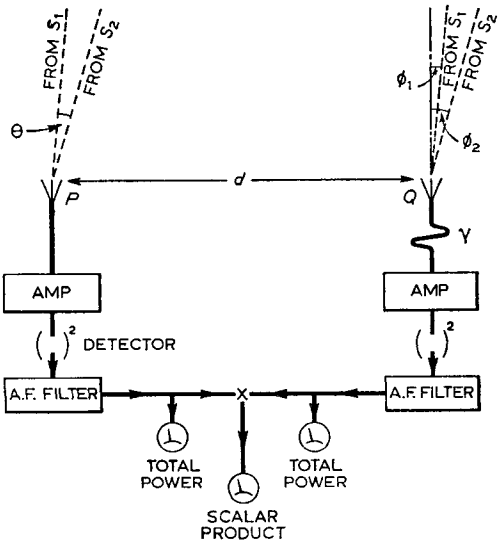
Later in this chapter we consider harmonic interferometers. The first, for spatial frequencies, is a somewhat deceptive instrument for, though at first it may seem to be possessed of very remarkable properties of resolution, on close inspection it is seen to be based on a similar fallacy to the nth power recorder system described earlier in this chapter. It is included here in order to illustrate the possibilities and pitfalls of an unorthodox

technique. Finally we consider a harmonic interferometer which does work, this time for spectral frequencies, correlating a signal with any small amount of its second harmonic which may fall upon the aerial.

It is important to note that any interferometer employing a non-linear device prior to cross multiplication, inevitably exhibits a poor signal to noise ratio in the presence of signals which are weaker than the background noise. This is a result of the introduction in the non-linear element of beats between all the Fourier components which are present in the pass band; the coefficients of the incoherent noise terms predominate in the output.

THE INTENSITY INTERFEROMETER

In the intensity interferometer a separate square law detector is employed in each limb of the system. The outputs of these detectors contain the cross beats between all the elementary radiators in the source which are selected, prior to detection, in



(a)

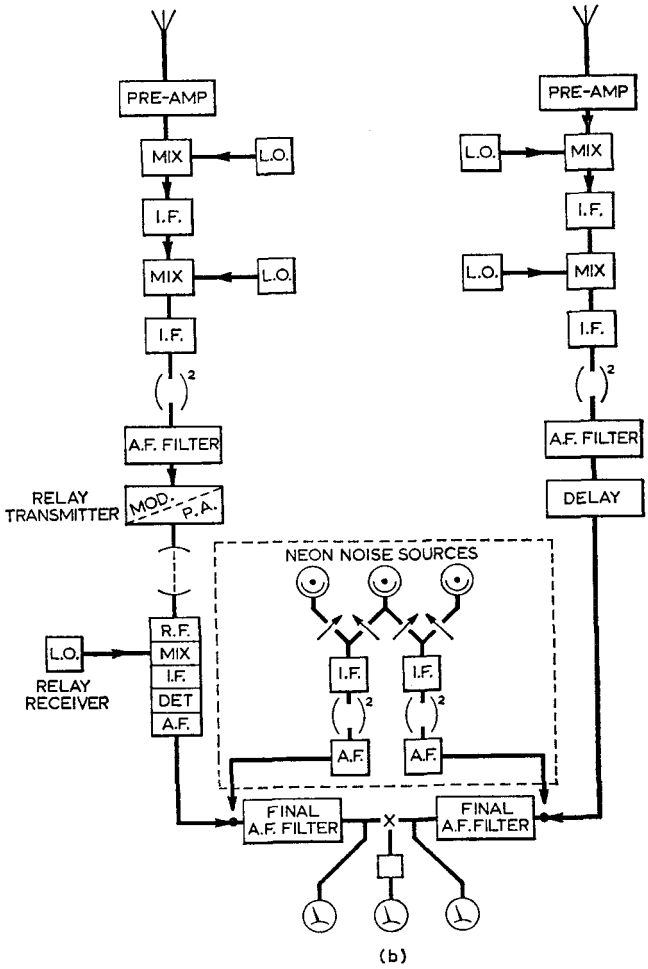


Fig. 31. The intensity interferometer

(a) basic diagram showing ray paths

(b) basic block diagram showing synthetic source used for calibration. Note that the local oscillators are mutually incoherent

the radio frequency bands of the two channels. The two radio frequency bands are arranged to be identical, and hence, in the detector outputs, the low frequency beat components due to the source have identical frequency spectra. The phase angle between corresponding components in the two channels is dependent upon the source brightness distribution function $S(\phi - \theta)$ and the baseline separating the two aerials, x measured in terms of the actual wavelength of the received radiation. The remainder of the apparatus consists essentially of a low frequency interferometer to measure the correlation of the two spectra; a monitor system for determining the total power in each channel and a synthetic noise source for calibrating the interferometer system.

The operation of the equipment is such that entirely separate local oscillators may be used, and the output of one channel may be transmitted to the mixing section via an incoherent radio link, in fact the outputs of the two receivers may conceivably be recorded and fed into the interferometer, or correlator section, at a later date. The basic diagram of the system is depicted in Fig. 31(a) and a fuller block diagram of the original Jodrell Bank instrument is given in Fig. 31(b).

If the radiation received by the two aerials is perfectly monochromatic the intensity interferometer gives no effect. Unlike the Michelson interferometer it requires more than one frequency to be emitted by the source in order that these frequencies may beat together in the non-linear elements at each receiver. For this reason the usual interferometry criterion of point-to-point correspondence must be replaced by a principle of point-pair to point-pair correspondence, where the frequency of each point in any pair differs from that of its associate by an incremental frequency which is within the pass-band of the audio filters of the receivers.

In order to illustrate the basic difference between this system and the Michelson interferometer we shall give a simple analysis following the same pattern as that used earlier.

Consider two points S_1 and S_2 located in the source at ϕ_1 and ϕ_2 giving rise to signals $\varepsilon_1 \sin(\omega_1 t + \psi_1)$ and $\varepsilon_2 \sin(\omega_2 t + \psi_2)$ respectively, where the phase angles refer to some instant at the point P. Then the signal from the points S_1 and S_2 (Fig. 31(a)) via aerial P to the receiver at P

$$= \frac{A_1}{\eta_1}\{\varepsilon_1 \sin(\omega_1 t + \psi_1) + \varepsilon_2 \sin(\omega_2 t + \psi_2)\}.$$

Similarly the signal from points S_1 and S_2 via aerial Q to the receiver at Q

$$= \frac{A_2}{\eta_2}\Big\{\varepsilon_1 \sin\left[\omega_1 t + \frac{2\pi}{\lambda}(d \sin \phi_1 + \gamma) + \psi_1\right] + \varepsilon_2 \sin\left[\omega_2 t + \frac{2\pi}{\lambda}(d \sin \phi_2 + \gamma) + \psi_2\right]\Big\}$$

We shall pass over the effect of the action of the frequency changers for the moment and consider that these inputs are directly applied to the square law detectors in each channel as in Fig. 31(a).

Then the output of the square law detector at P

$$= K_1\{\varepsilon_1 \sin(\omega_1 t + \psi_1) + \varepsilon_2 \sin(\omega_2 t + \psi_2)\}^2,$$

and the corresponding signal at Q

$$= K_2\Big\{\varepsilon_1 \sin\left[\omega_1 t + \frac{2\pi}{\lambda}(d \sin \phi_1 + \gamma) + \psi_1\right] + \varepsilon_2 \sin\left[\omega_2 t + \frac{2\pi}{\lambda}(d \sin \phi_2 + \gamma) + \psi_2\right]\Big\}^2$$

These signals are applied to audio frequency filters having a pass-band of, say, 1000–2500 c/s. These filters reject all the components formed by the expansion of the last two expressions except those in $(\omega_1 - \omega_2)$.

Hence the output of the audio filter at P

$$= \varepsilon_1\varepsilon_2 K_1 \cos[(\omega_1 - \omega_2)t + (\psi_1 - \psi_2)],$$

and the output of the audio filter at Q

$$= \varepsilon_1\varepsilon_2 K_1 \cos\left[(\omega_1 - \omega_2)t + \frac{2\pi}{\lambda}(d \sin \phi_1 - d \sin \phi_2 + \gamma - \gamma) + \psi_1 - \psi_2\right]$$

$$= \varepsilon_1\varepsilon_2 K_2 \cos\left[(\omega_1 - \omega_2)t + \frac{2\pi d}{\lambda}(\sin \phi_1 - \sin \phi_2) + (\psi_1 - \psi_2)\right].$$

The audio signals corresponding to the expressions for the outputs of the filters at P and Q are relayed to the correlator unit or interferometer section where the mean level of each is separately recorded by the meters R_1 and R_2.

Output of $R_1 = \overline{K_1\varepsilon_1\varepsilon_2}$,

and output of $R_2 = \overline{K_2\varepsilon_1\varepsilon_2}$.

The signals are also fed into a linear multiplier system giving

an output proportional to the scalar product of the two signals. This is recorded on meter R_3,

$$R_3 = \varepsilon_1^2\varepsilon_2^2 K_1 K_2 \cos\frac{2\pi d}{\lambda}(\sin\phi_1 - \sin\phi_2).$$

Division of the output recorded on R_3 by the product of those on R_1 and R_2 thus yields the correlation coefficient

$$\rho^2 = \frac{\varepsilon_1^2\varepsilon_2^2 K_1 K_2 \cos\frac{2\pi d}{\lambda}(\sin\phi_1 - \sin\phi_2)}{\overline{K_1\varepsilon_1\varepsilon_2}\cdot\overline{K_2\varepsilon_1\varepsilon_2}}$$
$$= \cos\frac{2\pi d}{\lambda}(\sin\phi_1 - \sin\phi_2).$$

If S_1 and S_2 are close together we may replace $(\sin\phi_1 - \sin\phi_2)$ by $\theta\cos\phi$, where θ is the angular separation of the point sources when viewed from the earth. Thus $\rho^2 = \cos\frac{2\pi}{\lambda}\theta d\cos\phi$, and hence if the source is in transit normal to the baseline so that $\phi = 0$, ρ^2 reaches its first zero when

$$\frac{2\pi\theta d}{\lambda} = \frac{\pi}{2}, \text{ i.e. when } \theta = \frac{\lambda}{4d}.$$

The preceding analysis for the bi-chromatic radiation of two points may be extended to cover radiation in a confined white spectrum from a continuous strip of random oscillators.

If points A and B are moved to all points within the limits $\Theta/2$ and $-\Theta/2$ the output of the linear multiplier (for $\phi = 0$, $\cos\phi = 1$)

$$R_3 = \int_{-\Theta/2}^{\Theta/2}\int_{-\Theta/2}^{\Theta/2} \varepsilon_1^2\varepsilon_2^2 K_1 K_2 \cos\frac{2\pi d}{\lambda}(\theta_1 - \theta_2)\,d\theta_1 d\theta_2$$
$$= \int_{-\Theta/2}^{\Theta/2} \varepsilon_1^2\varepsilon_2^2 K_1 K_2 \frac{\lambda}{2\pi d}\left[\sin\frac{2\pi d}{\lambda}\left(\frac{\Theta}{2} - \theta_2\right) - \sin\frac{2\pi d}{\lambda}\left(-\frac{\Theta}{2} - \theta_2\right)\right]d\theta_2$$
$$= \frac{\varepsilon_1^2\varepsilon_2^2 K_1 K_2\lambda}{\pi d}\sin\left(\frac{2\pi d}{\lambda}\cdot\frac{\Theta}{2}\right)\int_{-\Theta/2}^{\Theta/2}\cos\left(-\frac{2\pi d}{\lambda}\theta_2\right)d\theta_2$$
$$= \frac{\varepsilon_1^2\varepsilon_2^2 K_1 K_2\lambda^2}{\pi^2 d^2}\sin^2\frac{\pi\Theta d}{\lambda}$$
$$= \varepsilon_1^2\varepsilon_2^2 K_1 K_2\Theta^2\frac{\sin^2(\pi\Theta d/\lambda)}{(\pi\Theta d/\lambda)^2}.$$

The integrated outputs of the intensity recorders are,

$$R_1 = \overline{K_1\varepsilon_1\varepsilon_2\Theta} \text{ and } R_2 = \overline{K_2\varepsilon_1\varepsilon_2\Theta}.$$

If we now divide the mean reading on R_3 by the product of those on R_1 and R_2 we have $^2 = \rho\dfrac{\sin^2(\pi\Theta d/\lambda)}{(\pi\Theta d/\lambda)^2}$.

This expression will be recognized as of the same form as the square of the Fourier transform of a uniformly illuminated slit, and, as with a polar diagram or diffraction pattern of intensity, it can never be less than zero. We see that the first minimum occurs at $\Theta = \lambda/d$ as with the Michelson interferometer, though here the second maximum is smaller and of positive sign.

In the equation for the output of the linear multiplier, R_3, it will be noted that the limits of the integrals are identical, for in each case we are summing the contributions of the beat frequency components over the whole source. The definite integral of the first integration was treated as a constant in the second integration and we may consider the expression as the product of two identical integrals each extending over the source. These integrals are each equal to the Fourier transform of the source distribution whether this be a rectangular strip as in the present analysis, or any general function. Thus for any source, the variation of the output with aerial spacing corresponds to the square of the Fourier transform of the brightness distribution across the source.

The path difference between the two channels equally affects both of the beating signals entering a single aerial and cancels out in the expression for the output of the audio filter. Though in a fuller analysis it is of course necessary to take into account the loss in correlation incurred by a path difference which is comparable with the wavelength of the audio output, a considerable path difference at radio frequencies has no effect. The output does not therefore exhibit fringes but only a smooth curve as shown in Fig. 32, which depicts a typical record showing the intensity records R_1 and R_2 immediately above the cross product record R_3.

From this and the remarks in the preceding paragraph, it will be seen that the equipment is not phase-sensitive. This is a serious disadvantage if it is applied to the study of source structure, but was of little consequence, and indeed an advantage, for the original purpose of the equipment, which was to measure, with a high resolving power, the approximate diameters of the most powerful sources.

The beat components at audio frequency may be transmitted over very considerable distances and this was the chief merit of the equipment as it was originally conceived and constructed. It then seemed very probable that the sources in Cygnus and Cassiopeia were of very small angular diameter and might not be resolved at spacings up to hundreds of kilometres. Further reference to the liquidation of γ in the analysis shows that changes of phase in the two remote parts of the equipment are

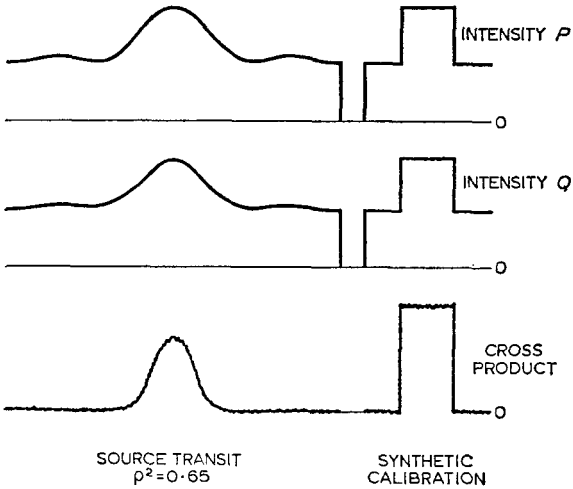


Fig. 32. The output of the intensity interferometer. The upper two traces show the outputs of the separate channels and the lower trace shows the cross product. The bump from the source on the left shows a correlation coefficient of 65 per cent by comparison with the unity correlation of the synthetic signal appearing as a rectangle on the right

of no consequence, and the oscillators serving each super-heterodyne receiver may be entirely incoherent. This is a decided advantage, for it is only necessary to transmit the audio output of one receiver by conventional methods without the need to synchronize the oscillators in each channel.

Again referring to the liquidation of γ, we see that changes in the ionosphere over each aerial, such as are experienced when there are intense scintillations, do not seriously affect the

results. This remarkable property is well born out in practice, even when the scintillations are very deep and the aerial spacing is over 10 km.

The interpretation of the phenomena that result from the absence of γ in the final outputs is as follows. The differential effect of path differences or phase shifts occurring prior to the non-linear system in each limb is to rotate the phase of the signals forming each beat frequency component by an almost equal amount, and the beat frequency itself is unaffected. Thus in the case of the inhomogeneity of the ionosphere, the signal paths from different parts of the source to aerial P are distributed over a very small region of the ionosphere, while those reaching aerial Q penetrate a separate small region. As the differential selection of the beat components occurs independently at P and Q the differences between the ionosphere above these two sites is of little consequence.

At baseline separations in excess of 100 km there is, however, another effect that must be considered. The difference in polarization caused by the Faraday effect in the ionosphere causes a reduction of the correlation. This must be taken into account in the assessment of ρ^2 if only one component of linear polarization is utilized.

The most serious limitation of this system of interferometry is a direct consequence of the term $\varepsilon_1^2\varepsilon_2^2\theta^2$ in the output of the linear multiplier. The output of the linear multiplier varies as the fourth power of the signal amplitude. There is not space in this account to derive a full expression for the final signal to noise ratio of the system, but it is evident that the noise background on the cross modulator is also dependent on the fourth power of the original uncorrelated noise. When the signals are initially equal to, or greater than, the background noise level, the accuracy is considerable, but a small diminution of the signal relative to the noise incurs a serious loss in the final reading accuracy. A signal which was initially only ten times down in power relative to the background noise is reduced to one hundredth in the record on R_3. For this reason the apparatus was entirely satisfactory for the initial measurements of the diameter of the two strongest sources, Cygnus and Cassiopeia, but was of little avail for the investigation of fainter sources. The advent of very large telescopes together with parametric amplifiers and masers would now enable the technique to be used again, if required, for the attainment of ultra high resolution over very long baselines. It is possible to synchronize

tape-recorders with sufficient accuracy to avoid the problems of the radio link on very extended baselines.

In the practical computation of the correlation, the use of the synthetic source shown above the cross modulator in Fig. 31(b) proved a considerable asset. This synthetic source enabled a fully correlated noise source to be applied to both low frequency channels together with uncorrelated components corresponding exactly to the conditions occurring at the transit of the source. By a direct comparison between the cross modulator

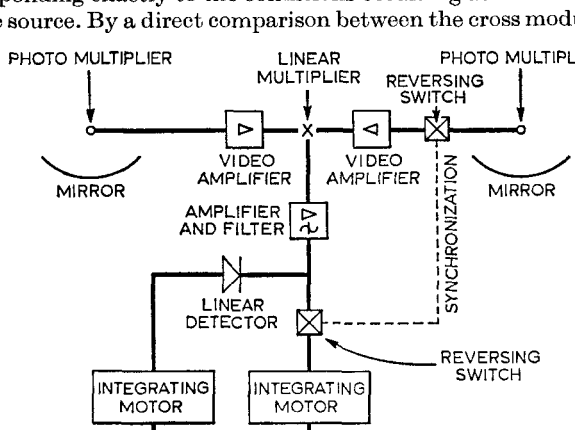


Fig. 33. Optical version of the intensity interferometer

output, R_3, during transit and the output of the cross modulator exhibited by the synthetic source for the same values of R_1 and R_2, the value of ρ^2 could be obtained almost immediately.

A later modification to the Jodrell Bank interferometer was the addition of facilities for operating on three adjacent frequencies. This enables the gradient of ρ^2 to be obtained at any point without the necessity of moving the aerials, and it was the method originally used to prove the existence of the second maximum in the transform of the Cygnus distribution. This was probably the first time that the variation of the parameter of frequency, rather than physical separation, was used in radio interferometry.

Radio astronomy has inherited many techniques from optical astronomy but the intensity interferometer has now found one of its most powerful applications in optical astronomy in the gifted hands of R. Hanbury Brown. The optical analogue was originally applied to measure the diameter of the star Sirius. It is not necessary to use optical surfaces of the usual standard and Hanbury Brown used searchlight mirrors to focus the light on to the photo multipliers, the video outputs of which were amplified and multiplied as shown in Fig. 33. A much larger and more sophisticated version of this instrument has been completed recently at Narrabri in Australia and has been used to measure the diameters of a number of hot stars.

HARMONIC INTERFEROMETERS

Harmonic interferometers provide interesting examples of the application of non-linear techniques and also illustrate their limitations. We will first consider an interferometer for spatial frequencies. As the principle of its operation is such that it may at first confound by its apparent simplicity, it has been included here to demonstrate the pitfalls of its superficial virtuosity; it does not work!

(1) *Spatial harmonics.* The resolving power of an interferometer depends upon the relative phase shift between signals from different parts of the source when compared between the two apertures. If some form of lever action can be introduced so that all relative phase angles are proportionately increased, it may seem that the resolving power is also increased. If we have a signal of the form $\varepsilon \sin(\omega t + \phi)$ we may double the phase angle simply by taking the second harmonic of the signal, thus obtaining a term $\varepsilon^2 \cos(2\omega t + 2\phi)$. If this is done for all the signals in one channel and they are then compared with their counterparts in the other channel it may at first appear that the resolving power of the instrument has been doubled.

The diagram in Fig. 34 illustrates such a technique. The signals in each channel are fed into frequency multipliers having selective circuits in their outputs whence any harmonic may be extracted and caused to interfere with the corresponding harmonic in the other channel. Consider a signal $\varepsilon \sin \omega t$ incident in the left-hand aerial, and a signal $\varepsilon \sin(\omega t + \phi)$ incident in the right-hand aerial.

If the two signals are amplified and passed into frequency

doublers, the second harmonic outputs in the two channels will be $\varepsilon^2 \cos 2\omega t$ and $\varepsilon^2 \cos 2(\omega t + \phi)$. The significant fact here is that the difference in phase between the two channels has been doubled, and hence the interference of the second harmonic produces a fringe pattern of twice the frequency:

$$\varepsilon^2 \cos 2\omega t \times \varepsilon^2 \cos 2(\omega t + \phi)$$
$$= \varepsilon^4(\tfrac{1}{2} \cos 2\phi + \tfrac{1}{2} \cos (4\omega t + 2\phi)).$$

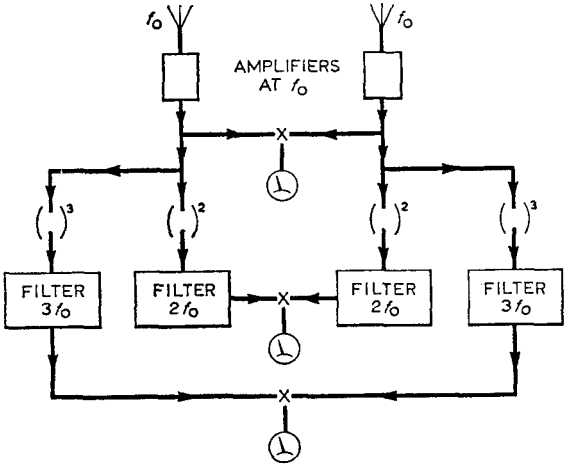


Fig. 34. No increase of resolving power is obtained with this harmonic interferometer

The second term is filtered out and the d.c. term applied to the recording meter whose deflection is therefore given by

$$R_2(\phi) = K_2\varepsilon^4 \cos 2\phi$$

which represents a cosine fringe pattern rotating at twice the frequency of the normal interferometer.

Similar treatment for the third harmonic yields a d.c. output to the recorder $R_3(\phi) = K_3\varepsilon^6 \cos 3\phi$.

The process could be repeated *ad lib* were it not for a severe deterioration in signal to noise ratio on weak signals.

From the above analysis it might appear that the system provides a simple way of achieving a high variable resolution at a fixed frequency and baseline. That this is not so is revealed by the application of the two-point test of resolution referred to earlier.

Consider now two signals incident on each aerial, giving

$$E_1 = \varepsilon_1 \sin \omega_1 t + \varepsilon_2 \sin \omega_2 t$$

in the left-hand channel, and

$$E_2 = \varepsilon_1 \sin (\omega_1 t + \phi_1) + \varepsilon_2 \sin (\omega_2 t + \phi_2)$$

in the right-hand channel.

In the frequency doublers these signals will be squared to give

$$E_1^2 = \varepsilon_1^2 \sin^2 \omega_1 t + \varepsilon_2^2 \sin^2 \omega_2 t + 2\varepsilon_1 \varepsilon_2 \sin \omega_1 t \sin \omega_2 t.$$

$$E_2^2 = \varepsilon_1^2 \sin^2 (\omega_1 t + \phi_1) + \varepsilon_2^2 \sin^2 (\omega_2 t + \phi_2) + 2\varepsilon_1 \varepsilon_2 \sin (\omega_1 t + \phi_1) \sin (\omega_2 t + \phi_2).$$

The second harmonic terms are now extracted, giving

$$f(E_1^2) = -\frac{\varepsilon_1^2}{2} \cos 2\omega_1 t - \frac{\varepsilon_2^2}{2} \cos 2\omega_2 t - \varepsilon_1 \varepsilon_2 \cos (\omega_1 + \omega_2) t.$$

$$f(E_2^2) = -\frac{\varepsilon_1^2}{2} \cos (2\omega_1 t + 2\phi_1) - \frac{\varepsilon_2^2}{2} \cos (2\omega_2 t + 2\phi_2) - \varepsilon_1 \varepsilon_2 \cos [(\omega_1 + \omega_2) t + \phi_1 + \phi_2].$$

It will be observed that a 'ghost' term in $(\omega_1 + \omega_2)t$ has been introduced; this term would not be present in a conventional interferometer operating on twice the baseline.

The multiplication of the two second harmonic signals yields,

$$f(E_1^2) \cdot f(E_2^2) = \frac{\varepsilon_1^4}{4} \cos 2\phi_1 + \frac{\varepsilon_2^4}{4} \cos 2\phi_2 + \frac{\varepsilon_1^2 \varepsilon_2^2}{2} \cos (\phi_1 + \phi_2) + \text{terms in r.f.}$$

This signal may now be compared to that from a conventional switched interferometer operating on twice the baseline; this would provide a fringe output proportional to

$$\varepsilon_3^2 \cos 2\phi_1 + \varepsilon_4^2 \cos 2\phi_2.$$

Though the terms in ϕ_1 and ϕ_2 reinforce and cancel as they go in and out of phase, due to the extension of the baseline to produce the visibility curve for a binary source, the term in $(\phi_1 + \phi_2)$ remains as a constant background and thus results in the first true zero appearing at the same baseline as in a simple interferometer; that is when, in this case, $\phi_1 - \phi_2 = 180°$. The term in $(\phi_1 + \phi_2)$ then cancels the other two, further extensions of the baseline causing the fringes to reappear. Now this corresponds to the behaviour of the visibility function in a simple interferometer, so from this point of view there is no improvement in resolution.

Similar treatments for the higher harmonics also show no improvement, and the reason for this is always the introduction of spurious cross beat terms in the two channels resulting from the non-linearity of the frequency multipliers.

In general it should be observed that the use of non-linear elements in interferometry does not increase the resolving power or enable additional information to be extracted about the shape and size of the source. Any system which is to result in an increase in resolution must achieve this by linear extension of the frequency or separation parameters.

(2) *Spectral harmonics.* If one sounds a musical note and then its octave, the two sounds are concordant and long ago Pythagoras showed that this correlation involved a simple factor of two. Nowadays we might call the octave the second harmonic and it may seem curious that, in the usual expression for correlation between mathematical functions, a fundamental wave and its second harmonic do not appear to be correlated. It is, of course, possible to construct a mathematical function which expresses the correlation, and it is also a simple matter to demonstrate this in practice. Figure 35 shows a technique once used by the author to investigate the possibility of coherent harmonics in various mechanisms and in particular in the radiation from the active Sun.

The spectrum was sampled in two bands, one of which was centred on a frequency of $2f_0$ which was exactly double the central frequency, f_0, of the other channel. The lower frequency signal was mixed with a local oscillator of frequency f_1 to give an intermediate frequency signal at $(f_0 - f_1)$ which was then amplified and fed into a square law frequency doubling circuit from which the second harmonic at $2(f_0 - f_1)$ was filtered and amplified. The signal output at the higher frequency, $2f_0$, was mixed with the same local oscillator after the latter had been fed through a continuously spinning phase-rotating network and a frequency doubling stage. The output of the mixer in this channel was amplified at the intermediate frequency $2f_0 - 2f_1$ which was commensurate with the final output of the signals in the first channel. The two bands of signals centred on $2(f_0 - f_1)$ were mixed in a balanced multiplier switched interferometer and the output of this was passed through a suitable time constant and fed into a pen recorder. Facilities were also provided for monitoring the total power in the two channels prior to their mutual interference. The block diagram in Fig. 35 shows an additional frequency conversion which was used in

the actual instrument. For simplicity the analysis in this paragraph has been confined to the signals which appear after the first two mixers in the diagram.

The equipment was extremely sensitive to a very small harmonic content in any signals reaching the input. When the

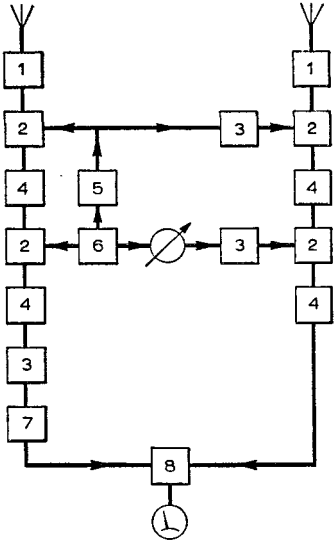


Fig. 35. Spectral harmonic interferometer for determining the presence of coherent harmonics

1. pre-amplifier
2. mixer
3. frequency doubler
4. I.F. amplifier
5. frequency multiplier $\times$ 10
6. local oscillator
7. filter tuned to second harmonic
8. switched interferometer employing balanced multiplication

The sign $\vec{\varnothing}$ represents a phase-rotating network

aerials at the inputs were pointed at the Sun, the system recorded a number of transient events showing coherent harmonics which appeared to be of solar origin both from the character of the events and from their temporal correlation with disturbances recorded at other observatories.

The simple grating. If a number of small aerials are distributed in a line and connected to a common receiver by feeders of equal length, the resulting aperture becomes the radio equivalent of the diffraction grating. The polar diagram of the

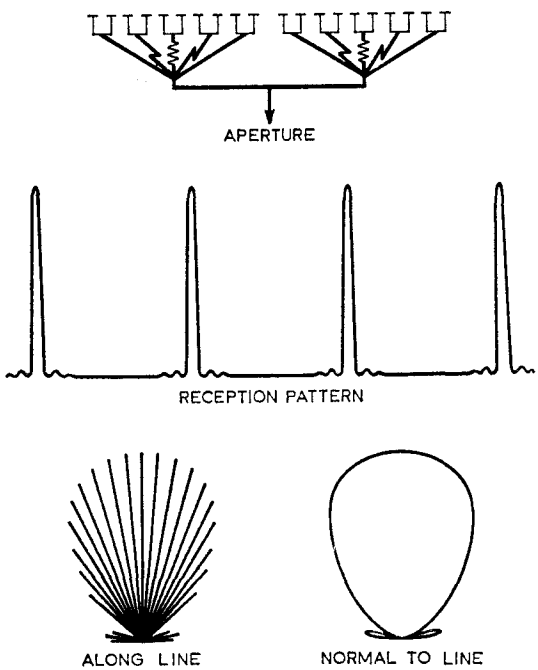


Fig. 36. Christiansen's radio analogue of the diffraction grating. The zig-zag lines in the upper diagram represent equal lengths of feeders

aperture is obtained by summing the Fourier components corresponding to the spacing of the elements. If n aerials are equally spaced, there are $(n - 1)$ cosine terms of equal periodicity and phase whose periodicity is determined by the separation between each element; $(n - 2)$ terms of twice this frequency; $(n - 3)$ terms of the third harmonic and in the

limit one term whose periodicity is that of the separation between the outermost elements. The summation of these terms yields a beam pattern which consists of narrow fan-shaped lobes having a width inversely proportional to the length of the grating (in wavelengths) and a periodicity determined by the spacing (in wavelengths) of the elements. A typical beam pattern such as would be obtained from an aperture consisting of 20 equally spaced and phased elements, is shown in Fig. 36.

The advantages of a multi-element interferometer were independently propounded by a number of workers, but the credit for the first practical realization of the system is due to Christiansen, who produced a 32 element system operating on 21 cm. The multiple-element interferometer is ideal for solar observations, for the solar radiation is subject to rapid and intense local variations which render the long term measurements of simple interferometers extremely unreliable.

The beam pattern of a simple multiple-element system is not suitable for radio star surveys or for the measurement of the structure of the radio sources because of the extreme ambiguity caused by the large number of beams. These may be considered as side lobes having powers equal to the central beam and even though the spacing of the elements may be chosen to permit only one beam at a time to scan a particular source, the remaining beams are liable to provide confusing information from small neighbouring sources.

There are two ways in which the overall beam pattern may be reduced in order to diminish the possibility of confusion. The first of these is to increase the bandwidth of the system to such an extent that the dispersion attentuates all but a few central lobes. This method is severely restricted by the liability of introducing man-made interference. The second method is based on the principle of concordant beams and will be discussed in a later section.

The ring telescope. A fascinating radio telescope for studying the Sun has been proposed by Paul Wild. It uses the principle that in a ring of telescopes all the Fourier components are available in the various azimuths between the various possible pairs of aerials, up to the spacing determined by the diameter of the ring.

The Sun is not usually directly overhead so that one must allow for the projection of the ring into the direction of the Sun and this implies that the ring should be an ellipse. The aerials are combined together at one of the focal points of the ellipse

and the beam then points in the direction from which the ellipse appears as a circle (Fig. 37).

This aerial system has a high resolving power, roughly determined by the minor axis of the ellipse, and a light gathering power corresponding to the total area of the aerials in the ring. If the aerials are simply connected in parallel at the focus the resultant beam shape is rather poor and could give rise to serious errors in the interpretation of the results. If one does not connect all the aerials together but considers all possible combinations of *pairs* of aerials in the ring one may take advantage of

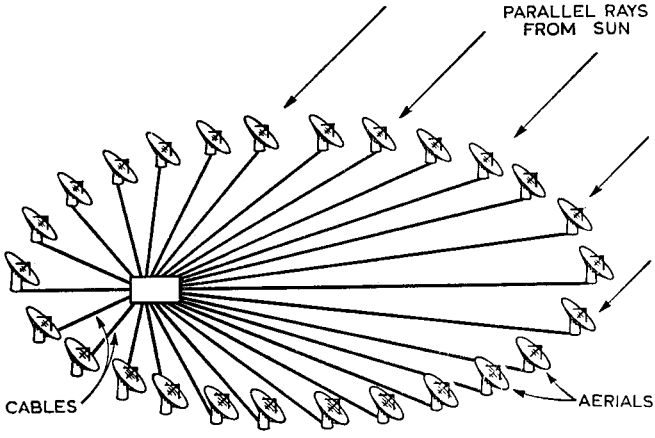


Fig. 37. The ring radio telescope

the interesting property referred to in the first paragraph. From the component aerials of the ring it is possible to obtain the outputs of a large number of interferometer systems having all possible orientations and spacings varying from those of adjoining aerials to aerials separated by the diameters of the ring. If the signal strength is sufficient, the response of the system may be arranged to correspond to that of one enormous dish, the size of the ring itself.

CONCORDANT BEAMS (UNFILLED APERTURES)

If we multiply linearly the voltage outputs of two aerial systems and observe the d.c. component in the output of the multiplier, we may trace, for each position of a point source, the modulus of the product of the electric fields which the source

produces in each of the aerial apertures. The d.c. output can only exist when the same signal is present in both channels, it therefore depends upon the concordance or correlation of the

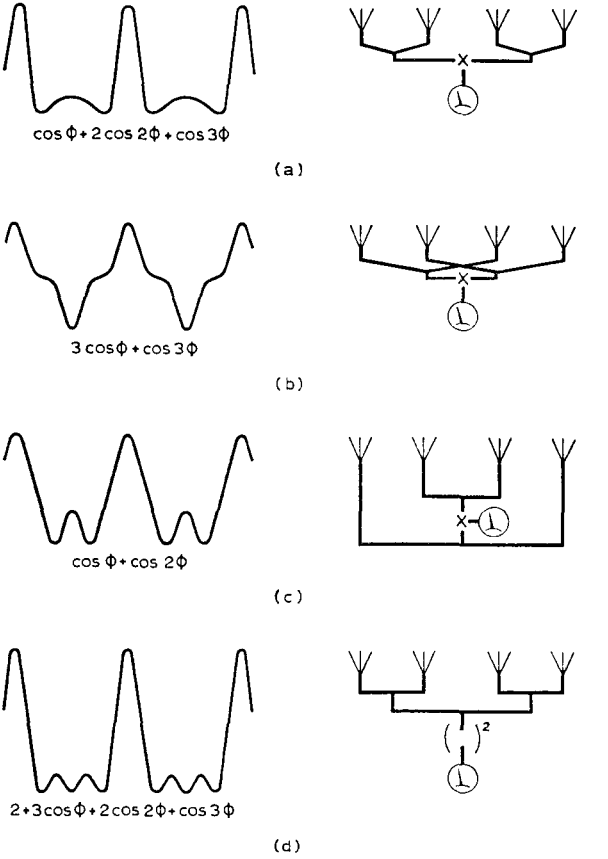


Fig. 38. Four aerials multiplied in three different ways (a), (b), (c), compared to the conventional addition and detection (d)

aerial beams, which also depend upon the aperture distributions. The most simple example of such a system is the phase-switched interferometer which we have already discussed. The effect of the switch is tantamount to observing the modulus of

the product of the signals from the two aerials, and in principle it is possible to dispense with the switch, and its associated detector and phase-sensitive rectifier, if we substitute a balanced hexode or similar linear multiplier from which the integrated direct current is recorded. Blum has used a variation in which the inputs are connected to opposite sides of a hybrid ring and the outputs from the two other terminals are separately detected and compared in a differential amplifier.

When the aerial system contains a large number of elements, the combination of beam patterns that may be derived from suitably combining the elements of the aperture is manifold. As an illustration of this, Fig. 38 shows the simple case of four aerials combined through a multiplier system, X, in three different ways, while the more usual polar diagram is also given for comparison. The base of the pattern does not necessarily coincide with the zero, the pattern in Fig. 38(b) for example, is symmetrical about the zero. In practice the position of the zero is also dependent upon the mutual coupling between the two systems. There seems little to recommend the reception patterns shown in Fig. 38. but we shall now discuss some cases in which the concordance of two beam patterns may be put to powerful advantage.

The product of a grating and a filled aperture. If a large conventional aerial is situated so that it fills the interval between the central elements of a diffraction grating, the outputs of the two systems may be linearly multiplied to give the result shown in Fig. 39. It will be seen that the effect is to select only the central lobe of the grating pattern giving a single knife edge beam. The power gain in the lobe is the product of the amplitude gain of the fan beam and the corresponding ordinate on the amplitude polar diagram of the central array.

The elimination of all but one of the beams of a multiple element interferometer has a serious attendant disadvantage. Though a fine beam may be produced the total observing time is seriously diminished. It is necessary to have a short integration time in the equipment in order to preserve the beam, but the number of independent observations is reduced to one. If, however, the central filled aperture is in the form of a large steerable radio telescope, the lost observations may largely be reclaimed whilst preserving the single fine beam and its attendant benefits from the avoidance of confusion. If the grating axis is East and West this may be achieved by steering the telescope into the region of each successive beam on the completion of each

transit of the combined beam. It is not essential for the tracking of the telescope to be accurate to within the final beam width. If the telescope is inaccurately steered on to a beam of the grating pattern the effect, in general, will only be to cause a small reduction in the output; the effect on the beam profile will

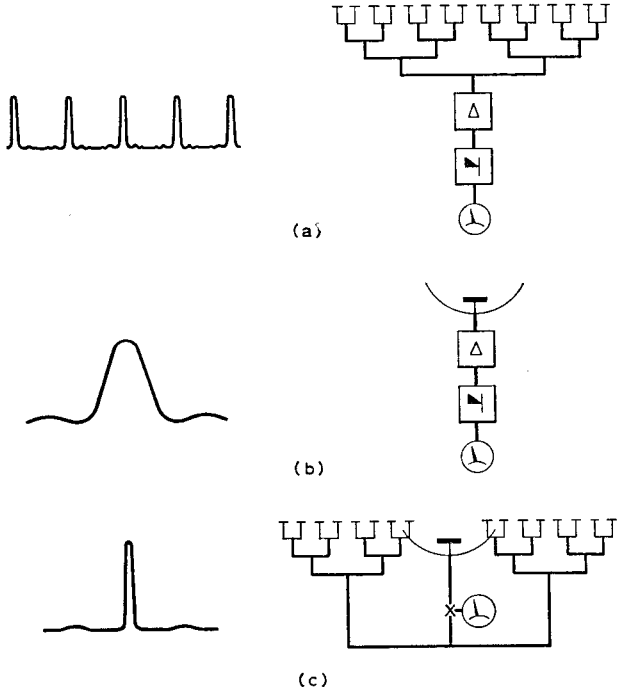


Fig. 39. The use of the principle of multiplication to produce a knife edge beam (c) from the directivity patterns of a grating (a) and a filled aperture (b)

be very small unless the error is considerable. In practice such a possibility of error could hardly arise, for it would infer that the tracking error of the telescope is of the same order as its normal beam width.

The crossed aerial. If two long narrow aerials having orthogonal axes are fed into a single receiver the resultant beam pattern is the sum of the two fan beams and will be reinforced

in the centre. If, however, the two aerials are multiplied together the resulting directivity pattern is in the form of a single pencil beam corresponding to the intersection of the two amplitude polar diagrams, Fig. 40(a). This beam pattern has the approximate form of the *amplitude* pattern of an aerial covering the whole square defined by the long aerials. The first side lobes therefore appear with negative 'intensity'. The effect of intersection with the side lobes may be largely eliminated by tapering the feeds of each aerial. The Mills cross aerial in Australia and the original design for the Benelux cross were aerials of this type.

Despite its high resolving power, the sensitivity of a cross system is very mediocre compared to that of the equivalent *conventional* aerial, for it intercepts only that part of the radiation which falls upon the long line of aerials. Nevertheless, the effective area of the giant new Australian cross will be greater than any of the present day steerable paraboloids. The increased resolving power makes the technique very attractive for surveying the sky in fine detail.

Cross systems may be used with two polarizations and, with difficulty, over a fairly wide range of wavelengths. In addition, if the lines of aerials are composed of separate dishes—as in the original Benelux design shown in Fig. 40(b)—a large number of outputs can be produced by connecting the dishes in a number of different ways. This increases the overall efficiency of the system, albeit at the expense of a considerable increase in electronic complexity.

The crossed grating. This is a variation of the cross aerial, sometimes called the Chris-Cross because it was first used by Christiansen in Australia. This technique multiplies the signals from two orthogonal lines of small aerials each spaced apart by a distance somewhat greater than their individual apertures so that each line of aerials is discontinuous and resembles a grating. The beam pattern obtained from the concordance of the polar diagrams is a 'bed of nails', a regular array of fine beams distributed over part of the sky. This beam pattern would suffer badly from confusion if applied to the radio stars but it is an excellent tool for studying the radio Sun.

MISCELLANEOUS RADIO TELESCOPES

Reflecting telescopes are usually of the Newtonian type, such as the Jodrell Bank telescopes, in which the rays are brought directly to the focus. Variations on this theme are Cassegrain,

Fig. 40. Cross aerial and derivation of beam pattern. The aerial system shown was an early design for the Benelux project

Gregorian and Coudé systems in which the converging rays from the main reflecting surface are intercepted by a secondary mirror which directs them down towards the base of the dish. With these arrangements the primary feed and associated equipment are easily accessible on the reflecting surface at the base of the bowl or on a short tower at the base. In either case the primary feed is well shielded from stray radiation from the ground.

Very large reflecting telescopes may be constructed by fashioning large bowls in suitable features of the Earth's surface such as the giant reflector in Puerto Rico. They may also be arranged on a flat surface by using the echelon principle. In this technique, shown in Fig. 41, a large number of flat or

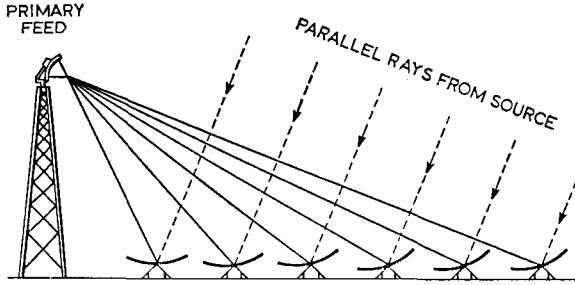


Fig. 41. An echelon telescope

slightly curved plates is arranged over a vast acreage to reflect the signals to a single primary feed aerial mounted upon a tower. The plates are servo controlled from a computer, so that the beam may be accurately focused and pointed over a considerable field of view.

Refracting telescopes on the Earth's surface would be impracticable for the wavelengths used in radio astronomy. In principle a refracting telescope can be made for very long wavelengths by placing a small dipole aerial on an artificial satellite in orbit above the F layers of the ionosphere. If the receiver is tuned to a frequency slightly higher than the ambient plasma frequency, the rays from the aerial are focused into the zenith by ionospheric refraction (Fig. 42). This focusing has the curious property that the signals are strongest near the edge of the resultant polar diagram, shown in Fig. 43, and very large gains are available in principle, corresponding at these wavelengths

to aerials far larger than any upon the Earth's surface. The focusing is very sensitive to both frequency and electron

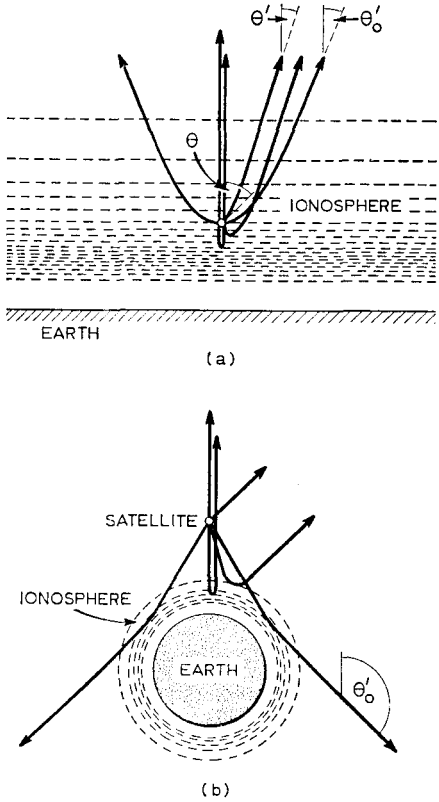


Fig. 42. Ionospheric focusing
(a) ray paths with aerial tuned slightly above the ambient plasma frequency
(b) ray paths far from the Earth for use in Earth occultation technique

density, and therefore to bandwidth and ionospheric irregularities. A better technique for achieving a high resolving power at frequencies of the order of a megacycle is to place the satellite in an orbit much further from the Earth and to survey the sky

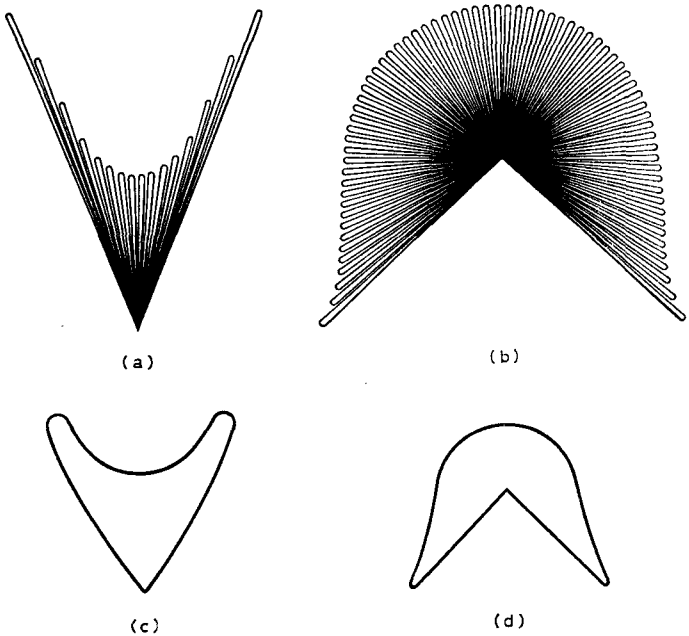


Fig. 43. (a) and (b) Polar diagrams corresponding to Fig. 42. The patterns (c) and (d) show the effect of a finite bandwidth and ionospheric inhomogeneity

by using the occultation of the Earth. If the frequency of the receiver is rapidly switched between two adjacent frequency bands and the output is synchronously detected and smoothed, the system has an equivalent polar diagram corresponding to a very narrow ring which sweeps across the sky as the satellite orbits the earth. This technique is far less sensitive to bandwidth and ionospheric irregularities.

APPENDIX

THE FOURIER TRANSFORM

Frequent reference has been made in these pages to Fourier components and the Fourier transform.

Fourier components are the elementary sinusoidal (or cosinusoidal) waves which contribute to a complex waveform. Any complex wave may be broken down into or built up from a number of simple waves of suitable frequency, amplitude and phase. Thus a complex musical sound can be synthesized by suitably sounding a number of simple tones, and conversely these tones may be extracted from the complex sound by the arrangement of suitable narrow band filters. A transient or isolated event such as a click or a bang can be represented by an infinity of closely packed and everlasting simple vibrations of suitably chosen amplitude. When a sharp click occurs these vibrations would all have the same phase at the instant of commencement but at no other time in past or future history, so that the waves only reinforce to form the sound at that moment when it actually happens. At all other times the integrated effect of the components is zero. The relationship which equivalently expresses the transient waveform as a spectrum of steady state frequency components is the Fourier transform. The relationship is reversible.

The same principles hold for many other processes including radio waves and electrical signals, the waveform of which (on a *time* axis) is related to the spectrum (on a *frequency* axis) by the Fourier transform. It is also found that one may apply the same treatment in the apparently very different domain of illuminated apertures and their associated directivity patterns; the amplitude and phase of the directivity pattern (in *angular* (sin θ) measure) is the Fourier transform of the amplitude and phase of the signal across the aperture (in *wavelength* measure).

The mathematical form of the transformation is given in complex form at the top of page 101. This relation may be reversed to give the field in the aperture in terms of the directivity pattern (in amplitude and phase):

$$a(x) = \int_{-\infty}^{\infty} F(\theta) e^{-2\pi i x \theta} d\theta$$

For a fuller descriptive treatment see: Jennison, R. C., *Fourier Transforms and Convolutions for the Experimentalist*, Pergamon, 1961.

INDEX

Aberration, 47
Absorption, 70
Acceleration, gravitational, 47
Aerial, 100
  , crossed, 49
  , grating, 144
  , isotropic, 101
  , power gain, 100
Allen, 85
Angular diameter, 1
Aperture, filled, 100
  , unfilled, 146
  , synthesis, 40, 105
Astronomical unit, 1, 41

Baade, W., 55, 77, 83, 85
Bandwidth limitation, 116
Barret, A. H., 76
Bay, 37
Beam pattern, 109
Beams, concordant, 146
Bed of nails, 150
Binary structure, 83
Black body temperature, 29
Blum, 148
Bolton, J. G., 113
Bracewell, R., 85
Brahe, Tycho, 53
Brault, 91
Bremsstrahlung, 26
Brightness distribution, 109
Burbidge, G. R., 86
Burke, B. F., 30

Calcium plage, 21, 22, 27
California Institute of Technology, 88
Carter, A., 126
Cassiopeia A, 9, 12, 46, 54, 62, 71, 72, 75, 76, 88
  spur, 56
Centaurus A, 85, Plate 10
Chris-Cross, 150
Christiansen, W. N., 21, 145, 150
Chromosphere, 27
Coherence, 121
Collapse, gravitational, 90, 91
Comets, 2
Concordant beams, 146
Condensations, 27
Confusion, 114
Convolution, 103
Conway, R. G., 62
Corpuscular streams, 23, 27
Cosmic epoch, 95
Cosmic rays, 23, 27, 51
Cosmological principles, 94
Cosmology, 91, 94, 95
Covington, A. E., 21
Crab Nebula, 30, 52, Plate 8
Crossed aerial, 149
Crossed grating, 150
Cyclotron radiation, 26, 47
Cygnus A, 4, 9, 12, 46, 52, 54, 72, 75, 77, Plate 9
  , loop, 57

Das Gupta, M. K., 55, 81, 82
Davies, R. D., 72, 75
Declination, 5
Degrees Kelvin (°K), 20, 29
Detector, 6
Deuterium, 76
Dewhirst, D. W., 54
Diameter, angular, 1
Dicke, R. H., 7, 29, 91, 98
Dipole, 6
Doppler broadening, 43
  radar, 43
  shift, 4, 15, 43, 48, 56, 65, 77
Drake, 36
Dyce, 35
Dynamic spectrum, 25, 27

Earth, occultation by, 152
Echelon telescope, 152
Einstein, Albert, 91, 96, 98
Electron density, 18
Energy, total, 49
Equivalence, principle of, 96
Evans, D. S., 86
Evolutionary cosmology, 95
Ewen, H. I., 36, 63

Faraday effect, 38, 39, 85, 137
Field, 72
Fitzgerald-Lorentz contraction, 47
Flare stars, 58
Flux density, 61
Forbidden lines, 78
Fourier component, 107, 144, Appendix
  transform, 49, 82, 101, 105, 109, 120, 135, Appendix
Fowler, 90
Franklin, 30
Fraunhofer diffraction, 101
  line, 70
Free-free transitions, 26
Fringes, 9, 109, 112
Fundamental observer, 95

Gain of aerial, 100
Galactic collision, 4, 77, 78, 79
Galaxy, normal, 45
  , nucleus of, 3
Ginsburg, V. L., 48
Gold, T., 90
Grating aerial, 144
Gravitational acceleration, 47
Gravitational collapse, 90, 91
Greenstein, J. L., 88
Gyro frequency, 22, 46

H I regions, 60
H II regions, 60
Hanbury Brown, R., 55, 81, 123, 125, 129, 139
Harmonic, coherent, 143
Harmonic interferometer, 129
  , (spatial), 139
  , (spectral), 142
Harvard University, 90
Hazard, C., 88
Henry, J., 76
Hewish, A., 21
Hey, J. S., 17, 22, 51
Hollinger, 85
Homodyne, 121, 122
Hoyle, F., 87, 90, 94, 97
Hubble, E., 98
  constant, 77
  law, 91
  time, 98
Hydrogen line, 13, 63

IC 443, 30, 57
Inertia, 96
Infra red, 30
Integration, 6
Intensity interferometer, 55, 81, 129
Interference pattern, 9
Interferometer, 9, 108
  , harmonic, 129
  , intensity, 55, 81, 129
  , Michelson stellar, 8
  , multi-element, 145, 148
  , non-linear, 129
  , phase-sensitive, 56, 83, 125
  , phase-switched, 11, 117
  , sea, 113
  , switched, 10
  , three-aerial, 56, 83
Ionospheric focusing, 152, 153
Ionospheric scintillations, 8
Isotropic aerial, 101

Jansky, Karl, 17, 45, 46
Jennison, R. C., 55, 82, 126, 129, 142
Jet Propulsion Laboratories, 44

Jodrell Bank, 5, 40, 42, 72, 81, 90, 132, 150, Plate 3, Plate 4
Jupiter, 31, 41, 44
, spectrum of, 33

Kellerman, K. I., 61
Kepler, J., 53
Kinematic acceleration, 47
Kinematic relativity, 98
Kwee, K. K., 68

Latham, V., 56, 83, 84, 124
Lequex, J., 84
Libration, 39
Light year, 3, 66
Lilley, A. E., 72
Lincoln Laboratories, 35, 42, 44
Lines, forbidden, 78
Little, 123, 126
Lobes, 9
, rotating, 124
Log N/log S curve, 91
Long, R. J., 61
Lovell, A. C. B., 58, 82
Lunar occultation, 88
radar, 37

M 31, 74
McClain, E. F., 72
McCrea, W. H., 97
Mach, Ernst (principle), 95
Magellanic clouds, 34
Magnetic field, 46, 48
Magnetohydrodynamic shock wave, 26
Maltby, P., 88
Malvern, 90
Mars, 36, 41
Matching, 105, 114
Mathews, T. A., 88
Maxwell, Alan, 28
Mayer, C. H., 84
Meeks, 76
Menzel, D., 36
Mercury, 34, 41
Meteors, 2
Michelson, A. A., 109
stellar interferometer, 8
visibility criterion, 112
Michigan University, 34, 36
Microwave early burst, 26
Mills, B. Y., 55, 81, 150
Mills cross, 150
Milne, E. A., 98
Minkowski, R., 55, 77, 83, 85
Moffet, A. T., 88
Moon, 29, 37
, occultation by, 88
Mössbauer, 91
Mount Palomar, 5, 55, 56, 88
Müller, C. A., 68
Multi-channel spectrometer, 13, 15
Multi-element interferometer, 145, 148
Murray, W. A. S., 38

Narrabri, 139
Naval Research Laboratory, 36, 84
Newton, Isaac, 96
Noise, 5
Non-linear interferometer, 129
Normal galaxies, 45

Occultation, by Earth, 152
, by Moon, 88
OH line, 76
Olbers' paradox, 95
Oort, J. H., 63, 70
Optical depth, 64
Optical thickness, 64
Orion, 59
Outbursts, solar, 25, Plate 6
Owens Valley, 85, Plate 5

Palmer, H. P., 88, 90, 125
Palomar, Mount, 5, 55, 56, 88

Parametric amplifier, 137
Parkes, Australia, 85, 88
Parsec, 66
Pawsey, J. L., 113
Payne-Scott, R., 123
Perfect cosmological principle, 95
Pettingill, G. H., 35
Phase of Fourier transform, 109, 123
  of visibility function, 83
  sensitive detector, 11
  switched interferometer, 116
Plage, calcium, 21, 22, 27
Planetary radar, 37
Planets, 41
Plasma frequency, 18
  oscillation, 26
Polar diagram, 9, 101
Polarization, 20, 21, 22, 25, 84
  of Jupiter, 34
  , Zeeman, 75
Power gain of aerial, 100
Principle of equivalence, 96
Pulsed radar, 43
Purcell, 63

Quasar, 62, 89, Plate 12
Quiet Sun, 5, 17

Radio link, 121
  star, 51
  telescope, 100
Rayleigh criterion, 103
Reber, Grote, 5, 46
Receiver stability, 6
Red shift, 4, 78, 89, 91
Red spot, 32
Refracting telescope, 152
Refraction in corona, 21
  in ionosphere, 152
Relativistic cyclotron frequency, 48, 51
  electrons, 51
  mass, 48, 51
Relativity, general, 98
  , kinematic, 98
Resolving power, 103
Right ascension, 5
Ring telescope, 145
Rosette Nebula, 60
Rotating lobes, 124
Rougoor, G. W., 70
Rowson, B., 56, 84
Royal Astronomical Society, 97
Ryle, Martin, Frontispiece, 7, 10, 40, 54, 94, 107, 109, 113, 117

Sagittarius, 46, 76
Sandage, A. R., 88, Plate 12
Satellite, artificial, 152
Saturn, 36, 44
Schmidt, 88
Schwarzchild, 91
Scintillations, 116, 136
Sea interferometer, 113
Seeger, Charles, 51
Selectivity, 61
Shklovsky, I. S., 48, 79, 86, 87, 89
Siderial day, 12
Slee, O. B., 21
Slowly-varying component, 21
Smoothing circuit, 6
Smith, F. G., 54, 81, 120, 129
Solar bursts, 26, 27
  corona, 18, 19
  limb brightening, 19
  outbursts, 25, Plate 6
  X-rays, 22
Spectra, 57, 61
Spectrometer, 13, 15, 24
Spectrum of Jupiter, 33
Stability, 115
Steady state cosmology, 95
Sudden Ionospheric Disturbance (S.I.D.), 23
Sunspots, 17, 21, 27

Supernova, 53
Supersynthesis, 107
Switched interferometer, 10
Synchrotron radiation, 26, 46, 51, 84, 86

Taurus A, 30, 46, 52, Plate 8
Telescope, radio, 100
, refracting, 152
, ring, 145
Thermal emission, 59
Thompson, A. R., 125
Three-aerial interferometer, 56, 83
Time constant, 6
Transistor, 6
Trifid nebula, 60
Twiss, R. Q., 126

Unfilled aperture, 146
Up-conversion, 123
Upper culmination, 12

Van Allen Belt, 23, 24
Van de Hulst, H. C., 63
Venus, 35, 41
Virgo A, 72, 85, 86, Plate 11
Visibility, 54, 109, 112
function, 82
Vitkevitch, V. V., 21
Vonberg, 7

Weinreb, 76
Westerhout, G., 68
Whipple, F. L., 58
Wild, J. P., 24, 25, 145
World picture, 95

X-rays, 23, 26

Yagi aerial, 8, 119

Zeeman splitting, 75

Plate 1. Spiral Nebula in Coma Berenices, seen edge on
(By courtesy of the Mount Wilson and Palomar Observatories)

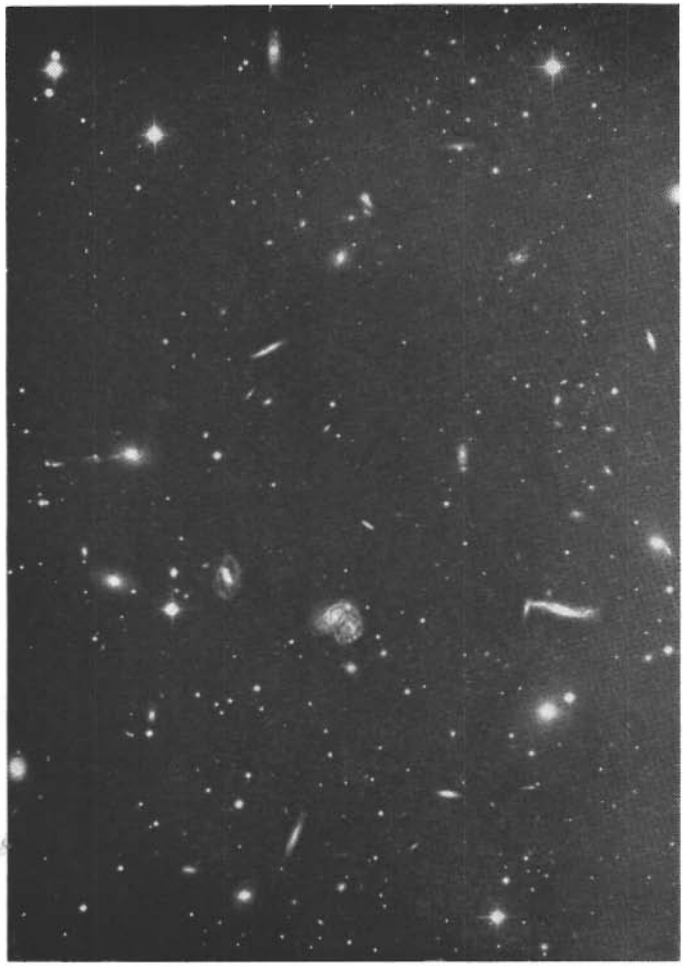


Plate 2. A cluster of extra-galactic nebulae in Hercules
(By courtesy of the Mount Wilson and Palomar Observatories)

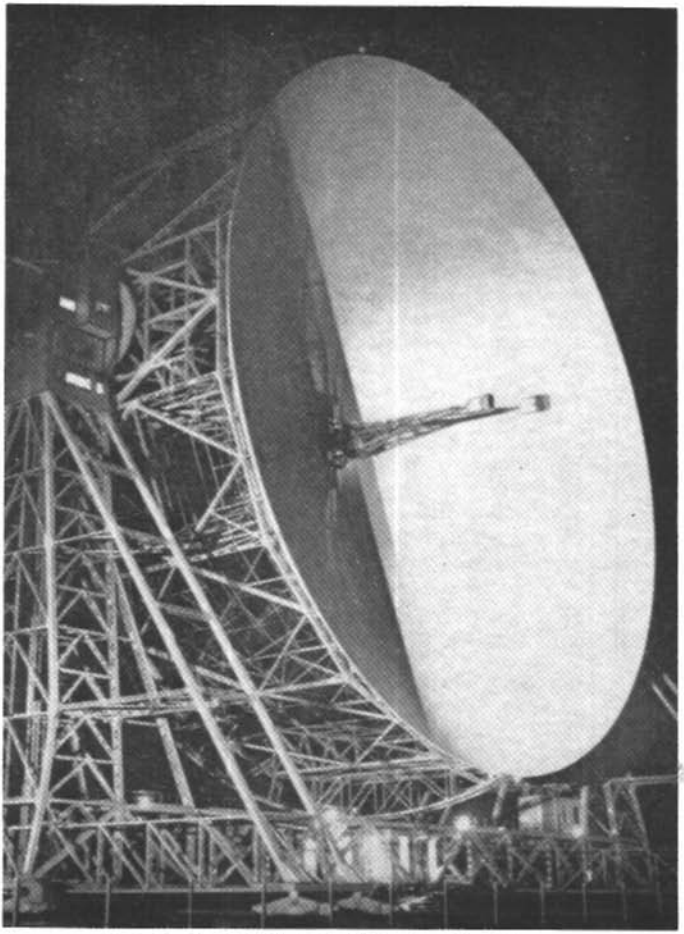


Plate 3. The 250 ft Jodrell Bank radio telescope

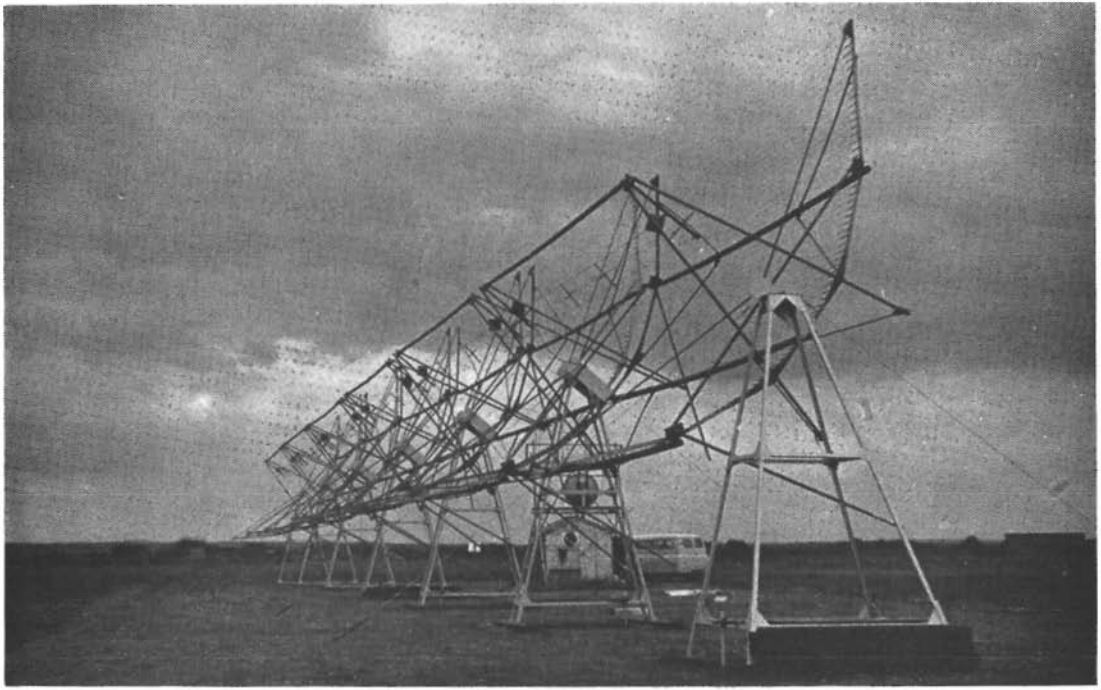


Plate 4. Interferometer outstation 100 miles from Jodrell Bank

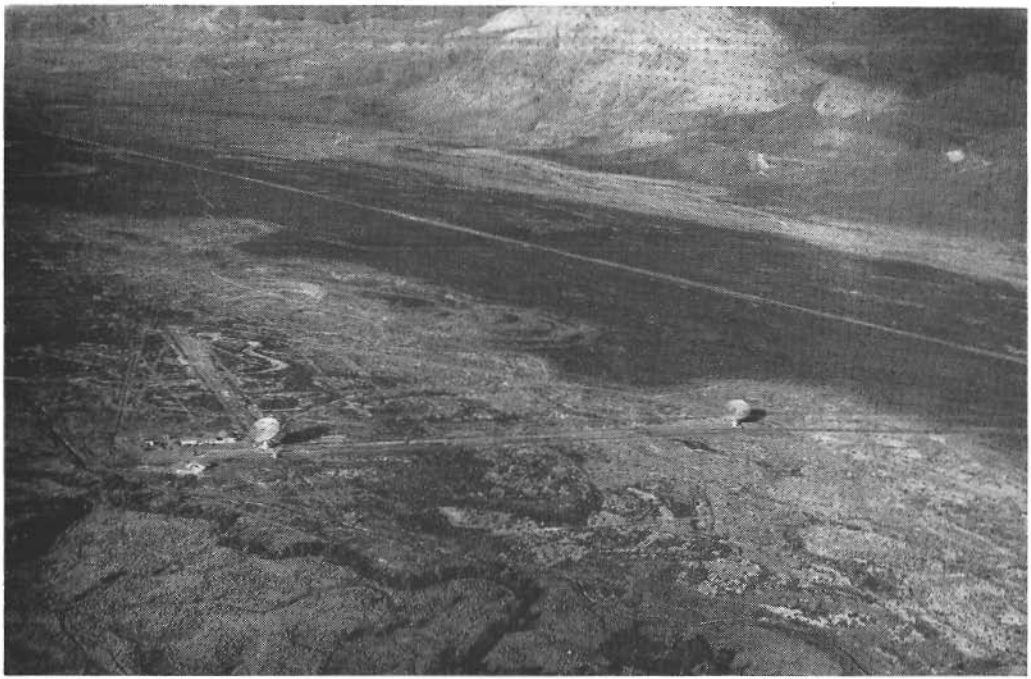


Plate 5. Two 90 ft paraboloids forming an interferometer, Owens Valley Observatory, California
(California Institute of Technology and U.S. Office of Naval Research)

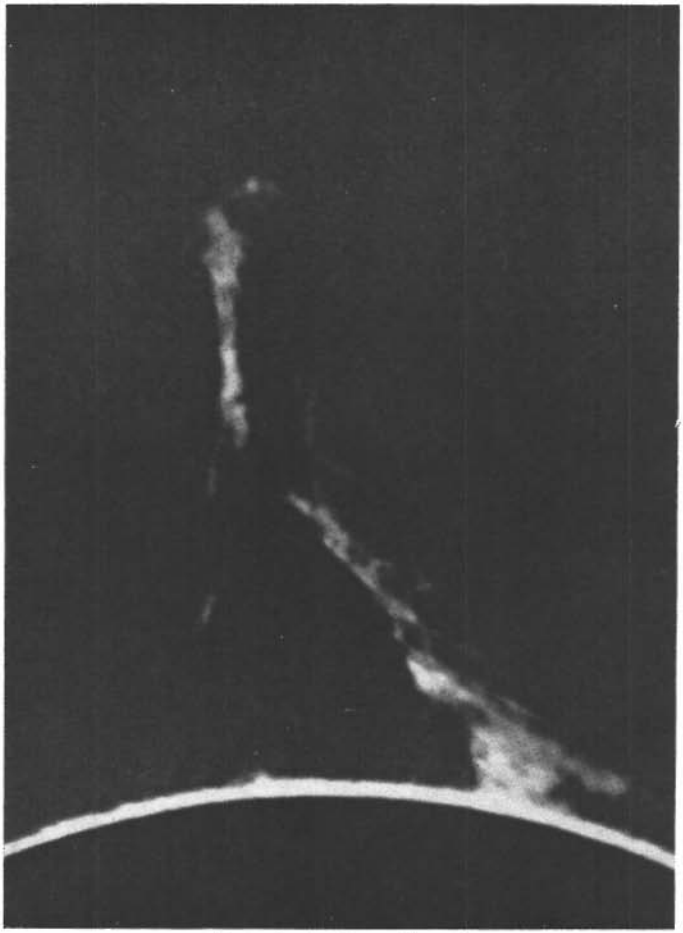


Plate 6. A solar outburst
(By courtesy of the Mount Wilson and Palomar Observatories)

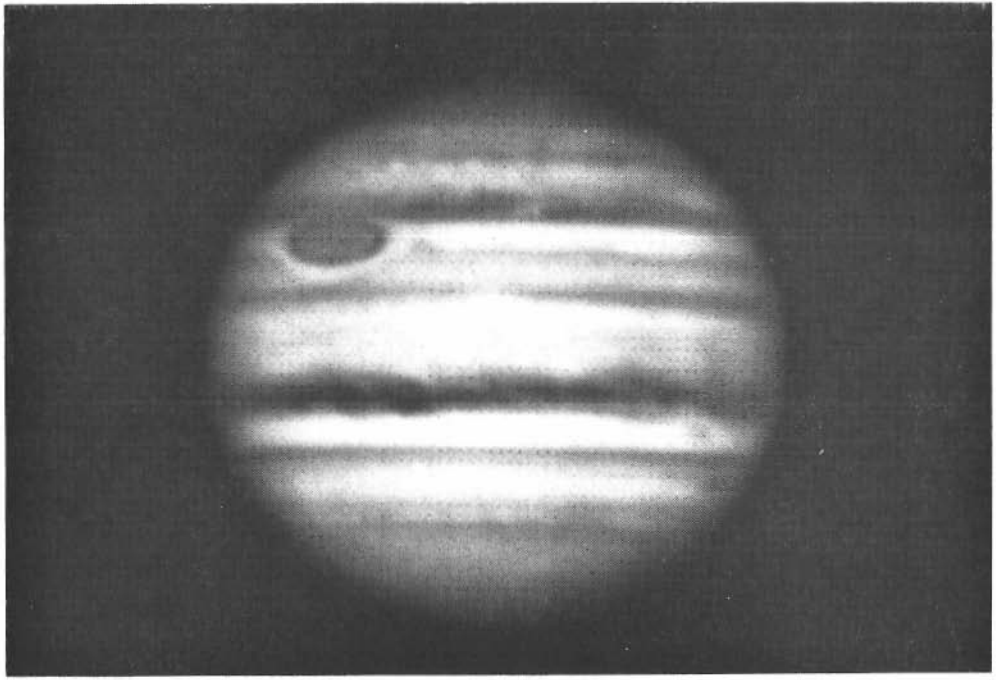


Plate 7. The planet Jupiter showing the giant Red Spot
(By courtesy of the Mount Wilson and Palomar Observatories)

Plate 8. The Crab Nebula in Taurus, remnants of the Supernova of 1054
(By courtesy of the Mount Wilson and Palomar Observatories)

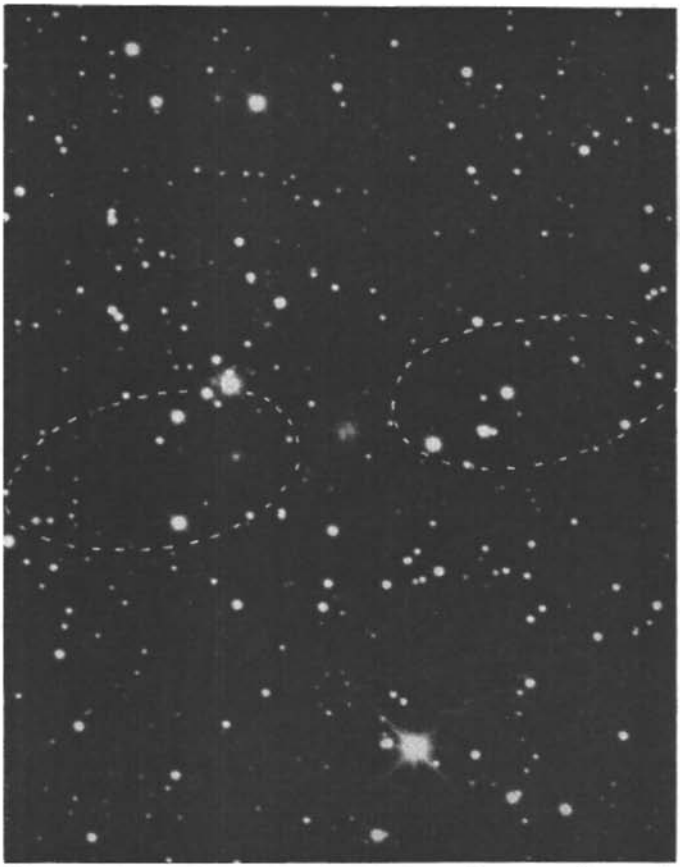


Plate 9. The nebula associated with the radio source Cygnus A (centre) showing the main components of the radio source (Superimposed on Lick Observatory Photograph)

Plate 10. The galaxy NGC 5128 in Centaurus showing the two central components of the radio source
(Superimposed on Mount Wilson and Palomar Observatories photograph)

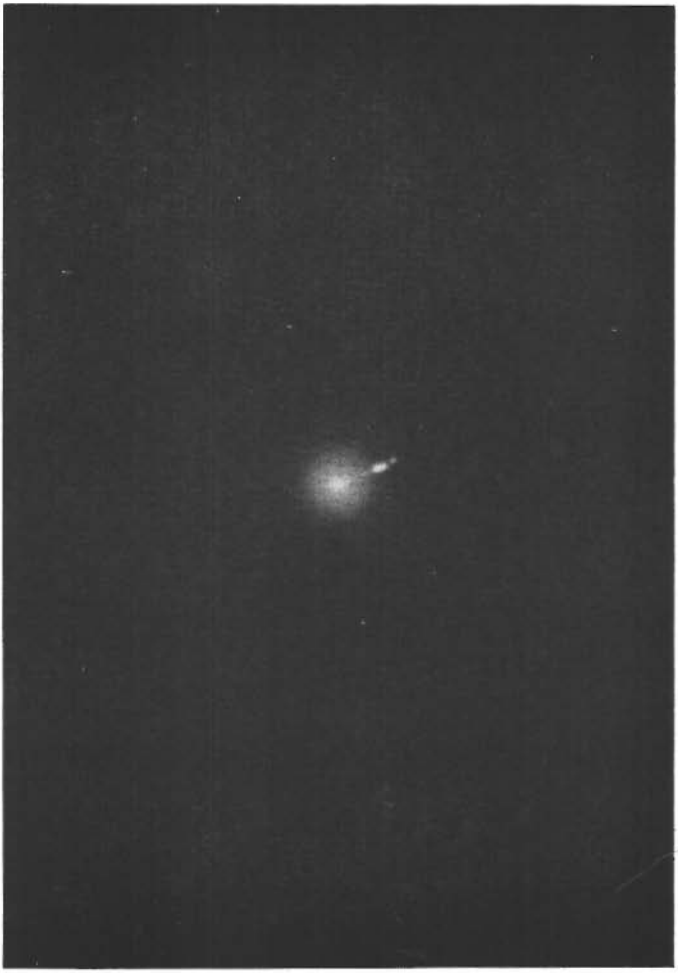


Plate 11. The galaxy associated with the radio source Virgo A. The blue jet pointing outwards from the centre emits polarized light consistent with synchrotron radiation. The radio source envelops the whole nebula. (Lick Observatory Photograph)

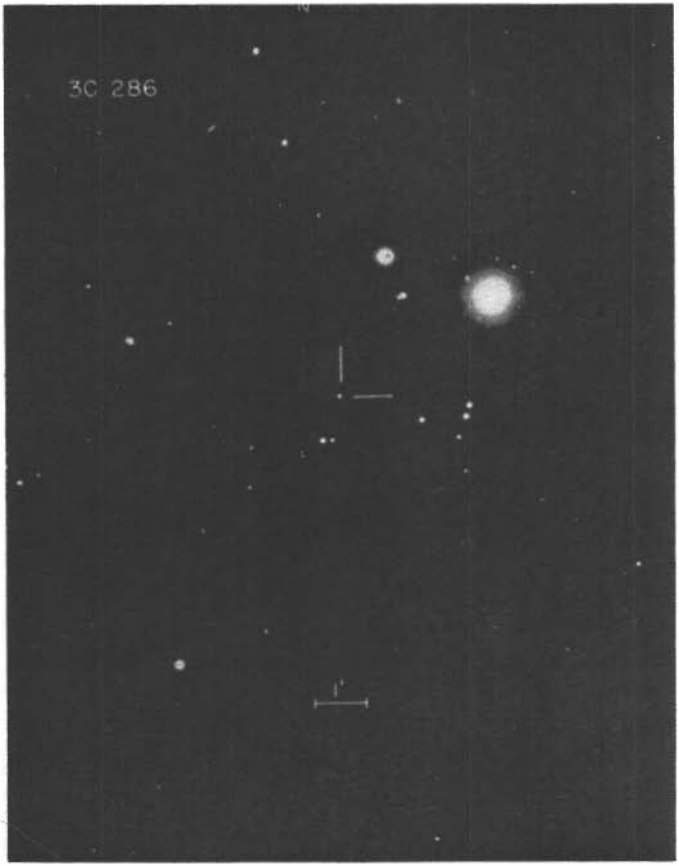


Plate 12. The 'Quasar' 3C 286 marked by the two white dashes (By courtesy of Alan Sandage, Mount Wilson and Palomar Observatories)

EDUCATION

Educated at the Grammar School, Grimsby, and at Hull Technical College, Dr. Jennison's studies for an engineering degree were interrupted in 1942 when he volunteered for aircrew duties in the Royal Air Force. He was grounded in 1945 as Ground Radar Technical Officer; resumed an academic career in 1947, this time in physics at Manchester University; graduated with an honours degree in 1950 and was awarded a Ph.D. degree in 1954 for work on the structure of radio sources.

The author first conducted experiments at Jodrell Bank while still an undergraduate. He was Research Assistant at Jodrell Bank 1950–52; MacKinnon Research Student of the Royal Society 1952–55; lecturer, then senior lecturer, in radio astronomy, Manchester University, 1955–66; currently Professor of Physical Electronics, University of Kent at Canterbury.

Dr. Jennison has been President of the Institution of Electronics since 1954, is the author of numerous learned papers and scientific articles on radio astronomy, electronics, space research and relativity, and of *Fourier Transforms and Convolutions for the Experimentalist*, Pergamon Press (1961).

RADIO ASTRONOMY

JENNISON

NEWNES